

Review of the International Dolphin Conservation Program Act Scientific DRAFT Report

by

Dr. Murdoch McAllister
London, United Kingdom

Review organized by the Center for Independent Experts, University of Miami

August 27, 2002

Executive Summary

1. The stock assessment work reported in the draft report and associated information that was reviewed represented a considerable improvement over the preliminary assessment reviewed in April 2002. The recommended revisions in statistical methods and additional analyses were carried out largely as recommended. The dolphin stock assessment group at the SWFSC should be commended for its willingness to consider the various recommendations provided in the April 2002 stock assessment review and the current one and its considerable efforts to revise its stock assessment analyses accordingly. However there are several aspects of the revised stock assessment as reported in the draft IDCPA scientific report and Wade (July 2002) that require correction and improvement.

2. The recommendation to carry out additional stock assessment evaluations of the potential for dolphin sets to induce unobserved mortality was implemented but the analyses had a number of shortcomings. It was recommended in the April review that a *calf* mortality covariate function be developed with annual unobserved calf mortality rate set to be a function of the annual number of sets * average annual number of animals set upon per set per year, divided by estimated population abundance. Instead, the mortality rates of all ages were made to be a function of only the annual number of dolphin sets divided by estimated dolphin abundance. The larger abundance estimates in the mid-1980s did not match well with the higher per capita sets in the mid-1980s; if calf mortality had been made a function of per capita number of sets as originally suggested, a better fit of the model to the data could potentially have been found. There was also a failure to report the significant positive relationship between mortality rate per capita (the estimated covariate parameter was positive and had a 95% probability interval that did not overlap with zero), and the (posterior) estimates of annual unreported mortality and r_{\max} both of which in this model are functions of dolphin sets per capita by year. It is recommended that further analyses be undertaken to more adequately evaluate the plausibility of dolphin sets causing unobserved mortality.

3. Bayes' factors were reported at 3.12 and 4.59 in favor of the simpler age-structured models over the unobserved mortality rate covariate models for ESD and NEOSD, respectively. However, these Bayes' factors and the results obtained were misinterpreted in the Draft Scientific Report to provide strong evidence against the mortality covariate models. Firstly, NMFS is not yet in a position to draw conclusions regarding the plausibility of the unobserved mortality rate covariate models because as mentioned above only a single relatively unrefined model for the unobserved mortality rate hypothesis has been considered. Only a single multiplicative statistical model for the covariate function was evaluated; an additive one was also suggested in the April 2002 review but remains to be evaluated. As mentioned above, while an all-ages mortality covariate function was evaluated, it is recommended that another one that models only calf mortality as a function of dolphin sets per capita to be considered instead, as originally recommended. Due to uncertainty over the effects of chase and encirclement, unobserved mortality could also be plausibly modeled as a function of the per capita number of chases rather than the per capita number of sets. A variety of other more refined model implementations will require rigorous evaluation before NMFS could find itself in a position to draw some modeling conclusions about the plausibility of the unobserved mortality hypothesis. In any case, any conclusions made about the plausibility of the unobserved mortality model need to take into account the various limitations in the analysis conducted. Secondly, the values obtained for Bayes' factors have been misinterpreted to indicate much stronger evidence than they actually provide (see next point). The values obtained instead provide only weak to moderate evidence in favor of the simpler models. Based on the current state of analysis, it could be more correct to conclude that when compared against a simpler model without the covariate, the unobserved mortality rate covariate models appeared to be slightly less likely given the data, and that refinements of this model and its data inputs could help to improve the fit of this model to the data and its relative plausibility.

4. Although Bayes' factors, instead of AIC, were computed as recommended, to evaluate the plausibility of alternative population dynamics model structures, the results for values of Bayes' factors larger than two were commonly misinterpreted in the Draft report, and many inappropriate conclusions were made based on these misinterpretations. Guidelines were provided in the draft report (p. 76) to indicate that values of Bayes factor between 3 and 12 were to be considered moderate support in favor of the model with higher probability (though these values could be arguably taken to indicate only weak evidence). Yet, findings for Bayes' factors as small as 2.12 and larger - up to the largest value obtained of 7.43 - were taken to indicate positive, strong or conclusive evidence. This was so for findings on the unobserved mortality covariate model and additional fishing mortality rate options for the generalized logistic models, as well as the 2 slope exponential models additionally fitted to the TVOD. The larger Bayes' factors for these models were misinterpreted as strong evidence, when only weak to moderate evidence was actually implied. It is recommended that all of these conclusions be revised to reflect appropriate interpretations of the Bayes' factors obtained.

5. In the April 2002 review, it was recommended that the TVOD series also be incorporated in the stock assessments, together with a methodology that would use them primarily to fill large gaps in the RV series, and be given no more weight in the estimations than the RV data. This recommendation was carried out only with the exponential models, and reservations were stated regarding the potential for the TVOD data to introduce bias in the stock assessment. Analytical work reported in this review indicates that it is unlikely that TVOD used as recommended will introduce bias, even if large abrupt changes in TVOD were to occur over a short time series. The main effects of using the bias-corrected TVOD indices should be to increase the statistical power of the tests of the plausibility of alternative stock assessment model structures. It is therefore recommended that all of the stock assessment analyses be conducted using both the RV and TVOD series in the manner identified in the April 2002 review.

Background (quoted and paraphrased from the statement of work provided by CIE)

The tuna purse-seine fishery has used the association between tuna and dolphins to fish in the eastern tropical Pacific Ocean for over five decades. Three stocks of dolphins were depleted by high historical levels of dolphin mortality in tuna purse-seine nets, with an estimated 4.9 million dolphins killed during the fourteen year period 1959-1972. After passage of the Marine Mammal Protection Act (MMPA) in 1972 and the increased use of fishing equipment and procedures designed to prevent dolphin deaths, mortality decreased during the late 1970s, 1980s and 1990s to levels that are generally considered biologically insignificant.

While changes in the fishery have greatly reduced the observed mortality of dolphins dramatically, the MMPA, as amended by the International Dolphin Conservation Program Act (IDCPA), requires that the National Marine Fisheries Service (NMFS) conduct research consisting of three years of population abundance surveys and stress studies to form the basis of a determination by the Secretary of Commerce regarding whether the "intentional deployment on, or encirclement of, dolphins by purse-seine nets is having a significant adverse impact on any depleted dolphin stock". The Secretary must make a final finding in this regard by December 31, 2002. It should be noted that this issue is controversial and particularly relevant to persons involved with NMFS, the US and non-US tuna industry, and environmental groups.

The topic of this review is the IDCPA Science Report that will be presented to the Secretary of Commerce, along with information obtained under IDCPA, and other relevant information to form the basis of the Secretary's final finding. The IDCPA Science Report is comprised of the results of all research activities required under section 304(a) of the MMPA, as amended by the IDCPA. Each major

component of this report has been separately considered in a series of independent peer reviews conducted by the Center for Independent Experts (CIE). These consist of: the Abundance Review (October 15-17, 2001), the Stress Review (February 4-6, 2002), the Ecosystem Review (March 6-8, 2002) and the Assessment Model Review (April 3-5, 2002).

This particular review is a follow-up of the assessment model review that set out to review the methods of analysis to estimate population growth rates and trends in abundance of eastern tropical Pacific dolphins (ETPD). Two different reviewers were appointed, Drs. Malcolm Haddon and Murdoch McAllister. These reviews were completed in May 2002 and since then, NMFS has had a chance to respond to the recommendations made in these reviews. This stock assessment review evaluates the most recent assessment modeling analysis, considering particularly whether the responses to the April review are adequate and sufficient. The Draft Scientific Report is also reviewed for its interpretations and conclusions regarding the most recent stock assessment results.

Description of Review Activities

The components of the research carried out specifically with respect to NMFS' incorporation of comments previously received in the Stock Assessment Review were reviewed. Comments received as a result of the CIE stock assessment review and explanations of how/why such comments were or were not incorporated into the report were reviewed.

The review activities included the following:

- Review of the draft IDCPA report and associated papers provided for this review
- Production of a written report of comments and recommendations of the draft report.
- Discuss via telephone, on August 16, 2002, with NMFS staff regarding the incorporation of comments and related questions.
- Revision of the written report based on those discussions.

In the body of the review, it was first considered whether the responses to the original review comments are sufficient and acceptable in a manner similar to the role filled by a journal editor when considering manuscripts revised in response to referee's comments. Following this, additional detailed comments are provided on the draft IDCPA Scientific Report and the stock assessment modeling work reported in Wade (July 2002).

Are the responses to the April 2002 review comments sufficient and acceptable?

A brief response to the recommendations of the stock assessment was summarized on p.53-55 of Wade (July 2002). The main recommendations of the reviewers are listed below and the responses to these are evaluated further below.

Comments from reviewer, Dr. Malcolm Haddon

Quoted and paraphrased from Haddon's conclusions and recommendations (p. 21):

"... there are a number of aspects of the models and data that should be taken into account when attempting to interpret the results of the modeling. These include:"

1. The limited number of data points available to which to fit the models. This will limit the number of parameters that can be estimated with confidence.
2. The inability of the data to provide information concerning any expected density dependent effects. This makes the density dependent terms effectively redundant at current population sizes.
3. That the dynamics of the different models be considered as they are expressed at current population densities, and whether each type is telling the same things for each species being considered.
4. That uncertainty around the estimates of unfished equilibrium dolphin population abundance and the related estimates of current depletion rates is poorly determined when only one model is included in the analysis. The different density dependent models need to be compared or both included in a single Bayesian analysis so as to include model uncertainty into the estimates.
5. That calf mortality has not been taken into account in the modeling nor in the estimates of total mortality. Both of these things (combined with the actual number of sets on dolphins) will have a marked effect on the modeling outcomes.
6. That the TVOD data be reconsidered to determine whether there are any sub-sets that could be taken from it and included in the model. These would need to be more homogenous in how the estimates were made than in the complete data set.
7. That the comparisons between the 1-slope and 2-slope models be treated with great caution because of the lack of data and lack of a mechanism for the regime shift (other than a correlation with oceanography)

The responses provided to these conclusions/recommendations, although quite brief, were on the whole acceptable. Whether the responses to all of these points are sufficient, however, requires some careful consideration and the issue of sufficiency of response will be taken up after listing the main recommendations from my April 2002 review.

Comments from reviewer, Dr. Murdoch McAllister

Recommendations for immediate implementation (paraphrased from p.2 of McAllister's April 2002 Review).

1. Use TVOD data also. It is recommended that within the stock assessment model, the annual rate of change of bias in the f_d indices of abundance be estimated using a linear model for trend-bias while assuming no trend-bias in f_i indices. This should help to improve the empirical basis for the current estimates of growth rate and the relatively poor ability to discriminate among alternative models because of large gaps in the data, especially the lack of abundance data from 1991-1997. If a rigorous approach to removing the trend-bias from the f_d indices is applied, the issue of assigning weights to the f_i and f_d becomes less contentious and empirically estimated variances based on fits of the models to the data could be applied. However, this should be done applying the constraint that the f_d indices are given no more weight in the estimation than the f_i indices (i.e., the stock assessment model variance for the f_d data should be constrained to be no less than that for the f_i data).

2. Age-structured modeling of mortality as a function of dolphin sets per capita. It is recommended that an additional age-structured model be developed that models incidental mortality rates on **one-year-old** calves as a function of the annual per capita index of exposure to dolphin purse-seine sets, the latter measured by the number of dolphin sets on the species \times the annual average number of animals caught or chased per set divided by the population abundance.

3. Use of Bayes' factors instead of AIC. When different population dynamics model structures were compared, the criterion for choice was AIC. Although AIC is widely accepted and considered a rigorous and objective criterion, it is difficult to interpret and leads to only one model being selected. It is recommended that Bayes' marginal posterior probabilities (i.e., Bayes' factors) be computed instead for each alternative model considered. These latter statistics are more rigorous and useful because unlike AIC they account for parameter uncertainty and have a probabilistic interpretation, e.g., the probability that each model is true when compared to the others, given the data. And unlike AIC, these probabilities can and should be applied in a decision analysis of alternative policy options.

4. Apply scenario-based approach. It is recommended that a scenario-based approach be applied to evaluate the plausibility of various factors that might have been impeding population recovery over the last few decades since the reported kills in dolphin sets were substantially decreased. Bayes' marginal posterior probabilities should be computed and presented with each alternative population dynamics model to indicate the relative credibility/plausibility of each given both the fi and fd data.

5. Evaluate plausibility of fishery-induced mortality model against other plausible models. It is recommended that particular attention be given to evaluating the plausibility of the fishery-induced calf-mortality model (Recommendation 2) against that of other models that do not implicate the tuna-purse-seine fishery as the chief cause for impeding ETPD recovery.

Five recommendations (not listed here) were also made for consideration in the near future, and responses by NMFS were generally agreeable with these.

The direct replies to these recommendations were very brief, mostly just "done". The key question addressed by this report is whether the responses taken as demonstrated in the report and supporting documents of subsequent work were sufficient to address the reviewers' recommendations.

To a considerable extent, Haddon's and McAllister's comments and recommendations address the same basic issues. They both make recommendations regarding: (1) The use of additional data to model and evaluate the plausibility of a positive relationship between the annual number of dolphin sets per capita and unobserved fishery-induced mortality, especially on calves (MH 5, MM 2,4,5); (2) the approaches applied to take into account stock assessment model uncertainty (MH 3, 4, 7, MM 3, 4, 5); and (3) the limitations to scientific inference about population dynamics imposed by the use of sparse and imprecise data for parameter estimation and model discrimination (MH 1, 2, 4, 6, 7, MM 1). The next section evaluates whether the responses by NMFS to address the main concerns and recommendations from the April 2002 review of the stock assessment are sufficient.

Do the responses by NMFS sufficiently address the main concerns of the April 2002 stock assessment review?

(1) Concerns and recommendations regarding the use of additional data to model interannual variation in unobserved fishery-induced mortality (MH 5, MM 2,4,5).

NMFS has responded appropriately to these concerns and recommendations. However, the extent of response so far does not enable any strong conclusions to be made yet about the plausibility of the hypothesis that dolphin sets cause considerable additional unobserved mortality. Analyses have been carried out to estimate the annual number of sets made upon, dolphins chased and dolphins set upon per set for the two key species, north eastern offshore spotted dolphins (NEOSD) and eastern spinner dolphins (ESD) and other species (Archer et al. July 2002). The stock assessment has been updated to incorporate estimates of the annual number of dolphins sets and to model unobserved mortality rate as a function of per capita dolphin sets per year. Comments and recommendations regarding the work done as described in Wade (July 2002) and the Draft Report of the Scientific Research Program are provided below.

Point 1: Executive Summary: Stock assessments p. 9. par. 3. line 8 (Draft Report of the Scientific Research Program). "However, models including this survival factor were shown to be statistically less probable." This resulted because the larger research vessel (RV) abundance observations in the mid-1980s did not match well with model predicted abundances given by the higher per capita number of dolphin sets in the mid-1980s. However, NMFS is not yet in a position to draw conclusions regarding the plausibility of the unobserved mortality rate covariate models because only a relatively crude first cut implementation of the unobserved mortality rate hypothesis has been considered.

Only a single multiplicative statistical model for the covariate function was evaluated; an additive one was also suggested in the April 2002 review but remains to be evaluated.

And while an all-ages mortality covariate function was evaluated, it is recommended that other plausible mortality effects be modeled, in particular a covariate function for calf mortality only, as originally recommended in the April 2002 review. If calf mortality had been made a function of per capita number of sets as originally suggested, a better fit of the model to the data could potentially have been found. Increased calf mortality in the mid-1980s could potentially result in some delayed impacts on total abundance due to delayed impacts on total reproduction. This approach could possibly do a better job of predicting the apparent declines in the RV data in the 1990s. It is therefore recommended that calf mortality instead of all-ages mortality be modeled as a function of dolphin sets per capita.

It was recommended in the April 2002 review that a *calf* mortality covariate function be developed with annual unobserved calf mortality a function of the annual number of sets * average annual number of animals set upon per set per year, divided by estimated population abundance. Instead, the mortality rates of all ages was made to be a function of only the annual number of dolphin sets divided by estimated dolphin abundance. Yet data on the average annual number of dolphins set upon per dolphin set exist for a good part of the modeled time series. It is therefore recommended that the dolphin sets per capita covariate be revised to incorporate this additional term as recommended in the April 2002 review.

Due to uncertainty over the effects of chase and encirclement, unobserved mortality could also be modeled as a function of the per capita number of chases rather than the per capita number of sets.

Quite obviously the subset of unobserved mortality models can become quite large if all different permutations of the recommended alternative model formulations are tried. These include calf versus adult mortality, additive versus multiplicative covariate function, covariate defined using annual estimates of dolphins set upon per dolphin set versus not using these data, and using per capita number of chases rather than only the per capita number of sets. If all permutations were tried, this would give rise to 16 different model runs. Rather than running all 16 permutations, a more efficient approach is to identify a base case run from first principles and then evaluate the goodness of fit of this base case relative to some selected alternative model formulations. For example, calf mortality, additive covariate, covariate defined using annual estimates of dolphins set upon per dolphin set, and using per capita number of dolphin sets could be considered to be the base case. This would then leave four additional alternatives to the base case, where the alternatives to each of the four input conditions are evaluated one at a time. If a large improvement in model fit was found, then the base case for the unobserved mortality rate covariate model could be updated to the alternative with the best fit.

In summary a variety of more refined model implementations will require rigorous evaluation before NMFS could find itself in a position to draw some modeling conclusions about the plausibility of the unobserved mortality hypothesis. In any case, any conclusions made about the plausibility of the unobserved mortality model need to be revised to account for the various limitations in the analysis conducted.

Stress and Other Fishery Effects, p. 23 bottom. "This leaves only very indirect and simple tests that could be done as part of the assessment modeling. As discussed below, those simple tests *did not support* the existence of substantial amounts of additional mortality as a function of the frequency of sets on dolphins, but important caveats apply to this conclusion." As mentioned below under the next section, the conclusion in the second statement is a misinterpretation of Bayes' factor. Also, it is recommended that a brief note about the nature of the caveats be given at this point in the text.

p. 26. par. 2. "our ability to resolve potential effects with the available data is limited, and such changes should not be *entirely dismissed from consideration*." This makes it seem as if conclusive or near conclusive results against the unobserved mortality covariate model were obtained. As mentioned below, none of the values for Bayes' factors obtained could be interpreted as giving conclusive or near conclusive results.

Point 2. Throughout the draft, important details of the results of the assessment that incorporate the effect of mortality as a function of sets per capita per year are given little attention and glossed over.

A positive relationship between annual mortality rate and the number of dolphin sets per capita was actually found (Wade, July 2002, Tables 16, 33) and the lower bound of the 95% probability interval for the "slope" parameter was larger than 0 for both ESD and NEOSD. This result lends some support for the hypothesis that there is a positive relationship between per capita number of sets and unobserved mortality. Yet this result and its implications are mentioned nowhere in either Wade's updated stock assessment report or the Draft Report of the Scientific Research Program. It is recommended that this statistically significant result be taken into account in assessing the plausibility of the hypothesis dolphin sets cause substantial unobserved mortality of dolphins.

Wade (July 2002) p. 18. par. 3. In the unobserved mortality covariate model, the unobserved rate of mortality, M_y and therefore r_{\max} change as a function of sets per capita. It is of interest to evaluate how r_{\max} changes over the time series and particularly for the latter part of the series when mortality rates could have increased due to the modeled increase in dolphin sets per capita. However, while the unobserved mortality rates and maximum growth rate (r_{\max}) were reported in the absence of dolphin sets for the unobserved mortality rate covariate model, the annual values for these terms were not reported. The posterior median estimates of the annual M_y and $r_{\max,y}$ for the full time series should be reported (i.e., median values and 95% PIs) and compared with the values for M_0 and r_{\max} obtained in the absence of dolphin sets. It is also recommended that plots of M_y versus sets/capita, and $r_{\max,y}$ versus sets/capita be plotted. Also, a retrospective analysis should be conducted in which the model estimated M_0 and r_{\max} under no fishery-induced unobserved mortality should be applied retrospectively to the full time series (otherwise using the posterior distribution of model parameters actually obtained) and the resulting population trajectory compared with the one with unobserved mortalities included.

It is recommended that Bayes' factors should be computed for the mortality rate covariate model and the non-covariate model using the TVOD and RV data and the recommended TVOD bias correction factor.

(2) *Concerns and recommendations regarding the approaches applied to take into account stock assessment model uncertainty (MH 3, 4, 7, MM 3, 4, 5).*

It was recommended in the April 2002 stock assessment review that Bayes' marginal posterior probabilities be computed instead of AIC for each alternative model considered. Bayes' marginal posteriors for alternative models give the probability of a particular model relative to the others considered given the data and allow for more than two alternative models to be compared. This distinction should be made clearer in Wade (July 2002). Although Bayes' factors were computed, it was not actually Bayes' factors that were recommended but rather marginal Bayes' posteriors for the alternative models. Bayes'

factors give instead the ratio of the probability of obtaining the (same) data, given the model for two alternative models. Bayes' factors are thus ratios of probabilities for the two alternative models. As such, they may be arguably less intuitive to interpret. Bayes' factors were mentioned in the April 2002 review recommendations only because they are a commonly applied analogue when there are two alternative models. Nonetheless, it is acceptable that Bayes' factors were applied instead of Bayes' marginal posteriors because they are analogous to the latter.

Thus, at least partly as recommended, Bayes' factors were computed and applied to evaluate the plausibility of alternative models. Bayes' factors were computed for pairs of alternative models in each of the stock assessments that used exponential, generalized logistic and age-structured models. Providing that Bayes' factors are applied in a systematic and sensible manner, and interpreted properly and consistently, then they can form a useful and intuitive basis for making empirical statements about the relative plausibility of alternative hypotheses about population dynamics.

On p. 76 of the Draft Report of the Scientific Research Program, guidelines to the interpretation of Bayes' Factors were provided based on the paper by Kass and Raftery (1994). It suggests the following:

A Bayes' factor between 1 and 3 is considered weak evidence that one model should be favored over another, 3-12 represents moderate evidence, 12-150 is strong evidence, and > 150 is considered decisive evidence in favor of a particular model (Kass and Raftery 1994).

These interpretations, though from an authoritative source, are themselves debatable and rather vague. Values of 1-3 would indicate a marginal posterior probability for the "favored" model of between 0.5 and 0.75. It should be acknowledged that when we obtain values like these or even much higher, we really cannot know which model is the "true model" or a truly better representation of nature. In fact, neither model might be a reasonable representation of the underlying processes generating the observations. If uncertainty in the data is considered, Bayes' factors of 1-3 could hardly be considered to be evidence in favor of one model or the other, since the associated marginal posterior values can easily reflect mere random chance in the data; if we happened to obtain another sample with the same number of observations or slightly modify the structure or covariate inputs of one of the models, Bayes' factors could easily swing in favor of the previously "dis-favored" model.

Values of 3-12 imply marginal posteriors of 0.75-0.92 in favor of one model and according to Wade (2002) are considered to provide "positive" evidence for the favored model. While Bayes' factor might be 12 and the posterior 0.92 for the favored model, there is still uncertainty over which model is truly better, albeit with less uncertainty than given by lower values for Bayes' factor. Even though we obtained a value of 92% for the favored model, we should keep in mind that the hypothesis represented by the other model could still be a more correct interpretation of nature. The lower probability for the disfavored model could result from random chance in the data, a poor choice of model formulation for the processes modeled or a low level of refinement in the inputs (e.g., covariates) to the model. Thus, even with a Bayes' factor of 12 or higher, a conclusion to discard the "disfavored" hypothesis could still be incorrect. Under a Bayes' factor of 12, it could be interpreted that given the fits of the models to the data, there is an 8% chance that, given the data, the disfavored model-hypothesis is actually correct.

When a particular model obtains a low probability or Bayes' factor, it should also be recognized that the hypotheses underlying a model might still be valid, only the model formulation for the hypothesis or some of the inputs to the model are poorly specified. Only after a variety of model formulations for the hypothesis have been evaluated and all found to have very low probability could it be concluded that the hypothesis is not well supported by the data, relative to other alternative hypotheses.

If we accept Karl Popper's axiom that hypotheses can never be proven to be true, only disproved, then we should focus primarily on evidence that indicates low plausibility. In any particular application, it should be asked, what is a marginally acceptable level of evidence (e.g., in terms of Bayes' factor) against one particular model before it can be concluded that it is not well supported by the data? This choice of a reference point value should be a function of the relative costs of being incorrect in hypothesis rejection and done with recognition of the extent and refinements applied in the modeling work developed for the statistical evaluation. This choice is analogous to the choice of alpha and beta probabilities in conventional hypothesis testing.

In conventional hypothesis testing in Frequentist statistics, we typically apply a value of alpha of 0.05, though this value could be smaller or larger, depending on the relative costs of type I and type II errors and the perceived limitations on sampling design. A value of alpha of 0.05 implies a willingness to accept a 5% risk of a type I error, i.e., falsely rejecting the null hypothesis when the null hypothesis is in fact correct. By analogy, if we were to desire a similarly low chance of incorrectly concluding that the data do not support a particular model, we might decide to consider that there is evidence against one of the two models when the marginal posterior probability for it drops below 5%. This would imply a Bayes' factor of 19 before a conclusion could be made suggesting that the data do not support a particular model. Even then, this does not provide conclusive evidence. Notice that Kass and Raftery (1994) only suggest that there is conclusive evidence with a Bayes' factor of > 150 (or posterior > 0.993).

For Bayes' factors less than 19, it could be argued that the evidence is still considered to be equivocal. However, as Kass and Raftery (1994) suggest, there could be intermediate categories implying that the data were slightly more consistent with one of the models. Given the above considerations, posterior values of 0.8-0.9 (or Bayes' factors of 4-9) could be considered to be equivocal and only *very slightly* in favor of one of the models. Posteriors between 0.9 and 0.95 or Bayes' factors of 9-19 could still be considered to be equivocal and only *slightly* in favor of one of the models. It should always be acknowledged that such results could fairly easily result from random chance or a low level of refinement in model and input data specification. Also, it should be kept in mind, that if we consider the two alternative models to have equal prior probability, (0.5 prior probability or Bayes' factor of 1 for each alternative model), a marginal posterior of 0.9 (Bayes' factor of 9) is a only relatively small shift away from the prior probability distribution. A posterior of >0.95 (Bayes' factor of 19) could be considered to be a reasonably substantial departure from the prior to suggest data less consistent with one of the models.

There is also a slight degree of inconsistency in the interpretation of probabilities in the stock assessment modeling work between what is considered "statistically significant" in terms of parameter estimation, and supportive and unequivocal evidence in terms of model selection. In parameter estimation, probability intervals of 95% are reported for each of the stock assessment modeling analyses. In many instances, the lower and upper bounds of the intervals are judged against key reference points, e.g., a value of 0, or intervals for a comparable estimated parameter. A standard of 95% PIs are applied here and reported in all of the tables in Wade (July 2002). Yet Bayes' factors from 2.12 up to 7.43 with associated marginal posteriors of 68% to 88% (assuming priors of 0.5) are taken as evidence in support of various model alternatives. Thus a much lower degree of rigor was applied as a basis for conclusions regarding model structure than for parameter estimates. It could be argued that at least as much rigor should be applied in model selection as in parameter estimation. It is recommended that the criterion for reference probabilities used in model selection be at least as rigorous as those applied in parameter estimation.

It should also be noted that due to the methods of posterior integration applied, e.g., importance sampling, the estimate of Bayes' factor or Bayes' marginal posteriors for alternative models might not be that precisely estimated. No diagnostics were reported to indicate the reliability of the Bayes' factors obtained. It is recommended that a diagnostic such as the coefficient of variation in the average of the importance

ratios (McAllister and Kirchner 2002) be computed to indicate the reliability of the Bayes' factors and marginal posteriors obtained for alternative models. Given the uncertainty in the estimates of Bayes' factors that are typically obtained from numerical integration, there is even further reason for caution in interpreting the values obtained.

The choice of a reference value for Bayes' factor requires close attention by NMFS regarding the issue of the evaluation of the plausibility of alternative models for population dynamics. While the interpretation of these categories of values for Bayes' factors in Kass and Raftery (1994) is debatable, it appears that the intent of the stock assessment was to apply these interpretations of values for Bayes' factors.

For ESD, the values obtained for Bayes' factors ranged between 1.13 and 5.11 (Wade July 2002, Table 17) and for NEOSD the values ranged between 1.02 and 7.43 (Wade July 2002, Table 34). In general, when Bayes' factors were found to be low, i.e., less than two, they were correctly interpreted as weak or inconclusive evidence. But values larger than 2 and up to the largest of 7.43 were generally misinterpreted as conclusive or strong evidence and were inconsistent with the general guidelines for interpretation given on p. 76. From the stock assessment analyses, a number of apparently definitive and strong conclusions were reached regarding model structure and the hypotheses about population dynamics implied based on values of Bayes' factor ranging between 2 and 7.43. Yet, according to Kass and Raftery (1994), these indicated results only weakly to moderately in favor of models with higher probability. Moreover, these results were in the lower half of the range of what Kass and Raftery (1994) regarded as moderate evidence (3-12), i.e. all less than the half-way point of 7.5. Taking a value of 7.43 this would imply a posterior probability of 0.88 for the more likely model. This is certainly not strong or conclusive evidence in favor of this model, but rather weak or moderate evidence. Therefore, it is recommended that the conclusions made based on the larger values of Bayes' factors, be revised to be made consistent with the guidelines to interpretation provided on p. 76 or some other defensible and consistent set of guidelines such as those provided above. A detailed list of the many places in the text requiring revision is provided immediately below.

p. 26. par. 3. "The assessment modeling does not support the possibility that there is substantial additional mortality as a simple function of the frequency of sets per year." This conclusion is inconsistent with the guide to interpretation of Bayes' factors on p. 76. A value of about 5.1 was obtained in favor of the simpler model and this suggests only moderate evidence in favor of the simpler model without the covariate, and not strong or decisive evidence against the mortality covariate model. This would indicate a marginal posterior probability of about 0.15 for the covariate model and 0.85 for the simpler model, and is by no means conclusive evidence. As mentioned above, the analyses conducted so far are only preliminary and consider only a single proposed model structure and approach to modeling the relationship between mortality and frequency of sets per year. A substantial variety of other plausible model structures for this mortality hypothesis could and should also be evaluated, as recommended above. This shortcoming of the analysis is, however, reasonably captured at the end of the paragraph. It should be taken into account that the covariate parameter was estimated to be positive and the 95% PI for it did not overlap with zero and this was the case for ESD and NEOSD. This lends some support for the notion that dolphin sets leads to unobserved dolphin mortality and it is important that this point is made in this document.

p. 80. last par. "Bayes' factor for these models indicated that the base run using actual reported fishery mortality, was strongly favored over the models with higher mortality (Bayes' factor is 2.43 for the 50% increase in fishery mortality scenario and 7.43 for the 100% increase scenario)." This interpretation is inconsistent with the general guidelines provided for the interpretation of Bayes' factors on p. 76. Values between 3 and 12 indicate only moderate evidence in favor of the model with higher probability, not *strong* evidence.

p. 82. "The model including per capita number of sets was estimated to be 4.6 times less probable than the model excluding the set data. The results do not support the possibility that there is substantial additional mortality occurring as a function of the number of sets on this stock per year, as estimated in this model." Again, there is a misinterpretation of Bayes' factor. Values between 3 and 12 indicate at most only moderate evidence for and against model alternatives. A value of 4.6 which would be analogous to a approximately 20% marginal posterior probability for the covariate model and 80% for the non-covariate model is only moderate evidence in slightly in favor of the simpler model. It could not be counted as strong or definitive evidence against the covariate model. I have often seen results of this type swing from being slightly in favor of one model to another from slight reparameterizations of the models and slight updates of just a few data points. Thus, in this case, a Bayes' factor of 4.6 is very far from providing conclusive evidence in favor of the simpler model. Moreover, in Wade (July 2002), the covariate parameter was estimated to be positive and the 95% probability intervals for the ESD and NEOSD did not overlap with zero. This in contrast supports the notion of a positive relationship between unobserved mortality and dolphin sets per capita. The stated interpretation that the results do not support the possibility that there is substantial additional mortality occurring as a function of the number of sets on this stock misinterprets the Bayes' factor, ignores the important result of finding a highly probable positive relationship between unobserved mortality and sets per capita, and thus requires substantial revision. This incorrect interpretation of the Bayes' factor results is made in several different places throughout the document and in all of these places this incorrect interpretation and statements associated with it require considerable revision.

p. 82. par. 2. "models without an effect of the number of sets per dolphin per year performed better than those with such an effect". As mentioned above, this interpretation is too strong and needs to be toned down. The terms "performed better" could be revised to "slightly better" but really, the results are equivocal since a positive relationship between sets per capita and unobserved mortality was found with high probability.

p. 83. par. 2. "the model including TVOD resulted in a Bayes' factor of 5.1 in favor of the 2-slope model ... Therefore, if one accepts the use of TVOD estimates as appropriate, these results lead to the conclusion that the growth rate of the eastern spinner dolphin stock changed from a positive to a negative rate in the late 1980s or early 1990s." This again is a misinterpretation of Bayes' factor. A value of 5.1 would indicate a probability of 0.15 for one model and 0.85 for another, which is definitely not conclusive evidence against one model and in favor of the alternative. As mentioned on p. 76, values between 3 and 12 indicate only moderate evidence for and against alternative models. Therefore, it is inappropriate to conclude on a Bayes' factor of 5.1 that growth rate has changed, only that there is weak to moderate evidence in favor of a change in growth rate. This conclusion therefore requires correction here and in the many other places that it is made throughout the document.

p. 85. par. 1 end. The Bayes' factor for the simpler model for ESD was about three and yet it was concluded that "These results do not support the possibility that there is substantial additional mortality occurring as a function of the number of sets on this stock per year, as estimated in this model." Again this is a misinterpretation of the relatively low value Bayes' factor which is only weak or inconclusive evidence. The conclusion also ignores the positive estimate of the covariate parameter between unobserved mortality and the per capita number of sets per year which also had a 95% PI that did not overlap with zero. This conclusion therefore requires revision throughout the text.

p. 85 par. 2. "models without an effect of the number of sets per dolphin per year *were better* than those with such an effect". This conclusion is not supported by the only borderline weak-moderate value for Bayes' factor of 3.

p. 88. par. 2 "However, the inclusion of this survival factor decreased the fit of these models to the data, which were then considered *improbable* according to their Bayes' factors." Again, this is a misinterpretation of the Bayes' factor values obtained of about three for ESD and four for NEOSD. Rather than suggesting improbable values, these indicate moderate probabilities of about 25% in favor for ESD and 20% in favor for NEOSD. These probabilities definitely do not suggest improbable results by any means and the conclusion requires revision to instead indicate that the results were inconclusive.

p. 88. par. 2. "Similarly, models with reported mortality scaled up by 50% and 100% performed *worse* than models with just reported mortality levels. That is, hypotheses about unobserved or unreported mortality at these levels (or below) *were rejected* by these analyses." Again, the Bayes' factors are misinterpreted and the interpretations are inconsistent with the general guidelines for interpretation provided on p. 76. The Bayes' factors of about 2.12 to 7.43 in favor of the simpler models should not lead to model rejection, as these provide only weak to moderate evidence in favor of the simpler models. The associated marginal posterior probabilities of about 0.33 to 0.12 for the unobserved mortality models should definitely not lead to their rejection and instead be taken either inconclusive or only very weak evidence in favor of the simpler models.

(3) *Responses to concerns and recommendations regarding the limitations to scientific inference about population dynamics imposed by the use of sparse and imprecise data for parameter estimation and model discrimination* (MH 1, 2, 4, 6, 7, MM 1).

The draft report and accompanying work (Wade July 2002) acknowledge the limitations imposed on statistical inference and model discrimination imposed by fitting models to sparse and imprecise data. This was done in the description of methods, and to some extent in the interpretation of results. The wide probability intervals obtained for example for depletion from carrying capacity in several of the estimations were appropriately attributed to having sparse abundance data and data over only a limited time period after the main depletions had taken place.

Some quantitative evaluations were also carried out to quantify the limitations on statistical inference imposed by the sparse data. For example, in a section on simple trend analysis, statistical power analysis is appropriately applied to indicate the statistical power in tests for different rates of increase in abundance. Power (the probability of correctly detecting a particular rate of increase) was found to be high ($\geq 95\%$) for increases of 3% or higher. It was only moderate (67%) for an increase of 2%.

An unfortunate consequence of using only the RV series of twelve data points over the last twenty-two years was that some of the key questions could not be adequately addressed. For example, when the simple 2 slope exponential model was fitted to the RV time series to test whether the population rate had changed in the latter part of the time series, results were suggestive of a change for ESD only (Bayes' factor = 2.12) and inconclusive for both species because of the sparseness and imprecision of the RV data. When slightly more sophisticated population dynamics models, e.g., the 2 K and 2 R_{max} generalized logistic models were applied to test for potential changes in carrying capacity and maximum growth rate, it is not surprising that only inconclusive results were obtained.

One recommendation made by both reviewers was for NMFS to include also (albeit in different ways) the tuna vessel observer data (TVOD) abundance indices to reduce the problems of sparse, imprecise and limited data (MH6, MM1). Haddon suggested a further analysis of the TVOD to produce an index that used more homogenous subsets of the TVOD on the assumption that none of the time series worked up so far could be adequately corrected for trend bias. The results could then benefit from having one or more of the bias factors removed. However, due to the successive changes in search methods over time (Lennert-Cody et al. 2001), the time series that might be deemed to be sufficiently homogenous might be too short to be of any use in stock assessment. The value-added in terms of, e.g., ability to identify a

change in growth rate in the time series, could be negligible from the addition of a single or two or more consecutive time series of 10 years or less because it would be less likely that a shorter time series would span the period over which a change occurred. The benefits of incorporating additional data would be diluted because additional parameters would need to be estimated for each homogenous segment added, i.e., a constant of proportionality and possibly also a trend bias correction factor. The latter might be necessary because there would still be the problem of uncertainty over whether the new TVOD time series could be considered to be free of time-trend bias.

Along the lines of Haddon's suggestion, Lennert-Cody et al. (2001) have already developed a revised TVOD index based on the half-normal line transect model that they consider to be less sensitive to changes in the probability of sighting animals on or very close to the track line. This can thereby reduce the potential bias introduced by the application of the simpler hazard rate model in the previous TVOD series. When compared against the hazard rate series, the half-normal model appears to remove an important source of time-trend bias, though several others still potentially remain.

Rather than deciding to leave out of stock assessment analyses commercial catch per unit effort (cpue) time series that could still include time trend biases when compared to actual trends in abundance (even after GLM analysis to remove potential biasing effects), cpue data are still included in many different stock assessments. In many cases, the potential for there to be time-trend biases is accepted but no attempts are made to correct for them (e.g., ICCAT 2000). However, recently there have been efforts to estimate within the stock assessment model, time trend bias in cpue data. One recent approach in Bayesian stock assessment is to model cpue catchability using annual random-walk parameters (Fournier et al. 1998; Parma 2001, 2002; Brooks et al. 2002; Porch 2002). Thereby potential time-trend biases in cpue observations can be estimated explicitly within the stock assessment model. This has been done when fishery independent data have and have not been available and it is perhaps most reliably done when fishery independent indices that are considered to be unbiased are also included in the stock assessment.

Along these lines, McAllister suggested in the April 2002 review using in the stock assessment the revised half-normal TVOD time series together with a simple bias-correction approach that incorporates a linear time trend bias correction function and gives no more weight in the estimation to the TVOD data than the RV data. An evaluation of the deviates between the two time series reveals a linear time trend in the deviates (Fig. 4 below). If the RV time series can be considered to be unbiased, then this implies that there exists an approximately linear time trend bias in the TVOD data that can be estimated within the stock assessment. In a manner analogous to the random walk models of catchability, the model can thereby predict values for the TVOD observations assuming that there may be a linear time trend bias in them. The introduction of bias in parameter estimation thereby becomes much less unlikely because the potential time-trend bias in these observations is explicitly taken into account. This can improve the empirical basis of the stock assessment. Many more observations will become available to improve the precision in parameter estimation and to facilitate the evaluation of the plausibility of alternative model structures. The TVOD data can thus take on a limited role in the stock assessment estimation by filling in the large gaps in the RV time series (i.e. 1991-1997). Note that no extrapolation would be entailed because the RV and TVOD data both start in the late 1970s and end in 2001. Moreover, there are three successive data points in the RV series at the end of the time series in 1998, 1999 and 2000 and these points should allow the trend in the TVOD data to be fitted accurately and precisely to that in the RV data.

NMFS responded cautiously by including the TVOD time series in estimations using the exponential model only. They thus demonstrated capability to follow the recommendation but a lack of willingness to implement it in all of the stock assessment modeling. The reason for this and my detailed response to the reservations expressed about the use of the TVOD series in the manner suggested in the April review are provided below.

Draft Report, p. 25. "The method proposed to remove bias from the TVOD series, which we applied as directed, may well have removed much of any existing bias, but it is very likely in our view that additional fundamental problems remain. For example, this method would not have removed an abrupt change in bias that occurred within a brief period. Such an abrupt change may have occurred in the early 1990's with the implementation of the AIDCP, or in the later 1980s with the shift to searching with bird radar and helicopters. Additionally, there are major inconsistencies with dolphin school size estimates from the TVOD in comparison to the ground-truthed research vessel estimates."

The authors identify an abrupt change in bias in the TVOD series that occurred within a brief period as a potential fundamental problem for the stock assessment. Though it is not explained precisely why it is a fundamental problem, this could presumably create an artificial estimated change in growth rate. However, for at least four reasons, providing that the recommended bias correction method is applied, an abrupt change in the TVOD time series is unlikely to present a fundamental problem for the stock assessment estimation.

(1) *The potential for the method not removing an abrupt change in TVOD over a short time period to create bias or an artificial estimated change in growth rate (of meaningful proportions) was found to be unlikely, after some simple spreadsheet modeling of alternative bias scenarios for the TVOD data.* Spreadsheet computations demonstrate that an abrupt change over a short period, either a sudden drop or increase in the bias trend in the TVOD, is unlikely to cause any meaningful bias (e.g., out by over 1%) in the estimated growth rates and that it is unlikely that this would create an artificially estimated change in growth rate around the time of the abrupt change. Take for example a hypothetical abrupt change in bias in the TVOD time series, let's say an artificial drop of -10% between 1990 and 1991. Ten percent is a relatively large annual change given the postulated reasons for an abrupt change and the observed deviations between the uncorrected TVOD and RV observations (Lennert-Cody et al. 2001). It is correct that this relatively large drop would remain in the "bias-corrected" TVOD series (See figures 1-3 below). However, the bias-corrected TVOD series would be centered over the unbiased RV time series. The overall estimated average rate of increase of the time series would be unbiased, even with the large drop in bias in 1991 because the slope of the TVOD series is adjusted to that of the RV series. The first (1980-1990) and latter (1991-2001) parts of the bias-corrected TVOD time series, however, would have slight positive biases in their estimated trends. For example, if the true rate of increase was 2.0% and the annual steady change in bias was -1%, the spreadsheet calculations show that the rate of increase in the TVOD series would be 2.5% for 1980-1990 and then 2.4% for 1991-2001 (Fig. 1a). The estimate for 1980-2001 from the bias-corrected TVOD is still unbiased at 2%. Because the stock assessment is fitted to both the RV and the bias-corrected TVOD series, the rate of increase averaged over the RV and bias-corrected TVOD data for the first and latter parts would be 2.3% and 2.2% (Fig. 1a). These values are hardly different from each other and the true simulated growth rate of 2%. The abrupt change in bias simulated thus did not result in any meaningful artificial change in the averaged growth rate after 1991.

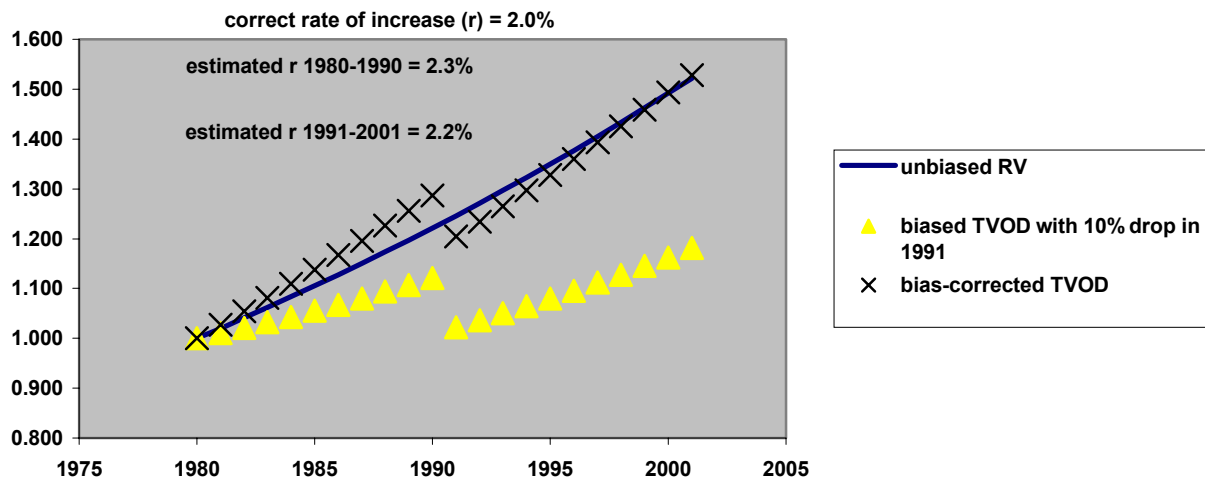
On the other hand, if there was an abrupt increase in bias in the TVOD by 10% in 1990, the averaged rates of increase for the first and latter parts would be 1.7% and 1.8% (Fig. 1b). Qualitatively similar results were obtained when the magnitude of the abrupt change in bias in 1991 was increased to 20% and also when the true growth rate was decreased to 1% (Figs. 2 and 3). It can be seen that the effect of an abrupt change in bias over a short period in the TVOD series on the estimated growth rates are negligible and that it is unlikely that an abrupt change in bias in the TVOD series could give rise to an artificial estimated change in growth rate of the amounts estimated using the actual RV and TVOD data.

In the stock assessment analysis, the probability of the data was computed under assumptions of no change versus a change in growth rates at some point in the time series. The model that was fitted assuming a change in growth rates was found to have up to about five times higher probability when the TVOD data were included. If there was truly no underlying change in growth rates yet there was a jump

in the bias in the RV data, the spreadsheet analyses demonstrate that no meaningful change at the time of the jump could be expected in the averaged trend in the RV and TVOD series. This implies that it would be unlikely for the TVOD series even with a large abrupt change in bias to create an artificially large Bayes' factor in favor of the model with a change in growth rate when none had actually occurred.

Figure 1. Deterministic trajectories of an unbiased RV time series together with a biased TVOD and bias-corrected TVOD time series. The true growth rate is 2.0%, the steady change in bias is -1% per year. (a) The drop in bias in 1991 is -10% . The estimated growth rate over the time series using the average of the TVOD and RV series was unbiased at 2.0%. The estimates for each of the two decades were slightly positively biased. Yet no meaningful artificial change in estimated growth rate was obtained in 1991, as in all of the other trials below. (b) The jump in bias in 1991 is $+10\%$. The estimated of growth rate over the time series using the average of the TVOD and RV series was unbiased at 2.0%. The estimates for each of the two decades were slightly negatively biased.

a.



b.

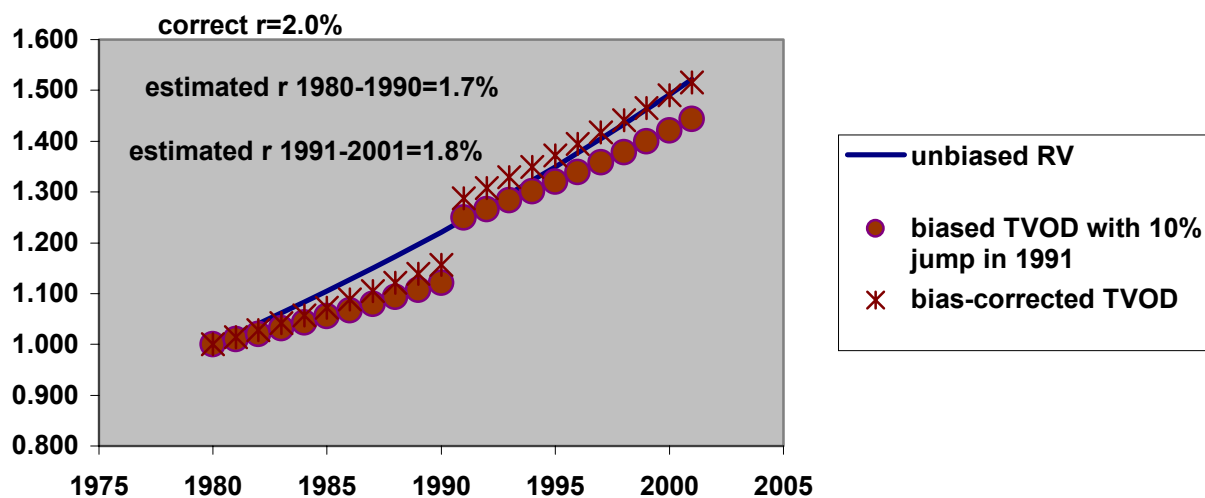


Figure 2. Deterministic trajectories of an unbiased RV time series together with a biased TVOD and bias-corrected TVOD time series. The true growth rate is 2.0%, the steady change in bias is -1% per year. (a) The drop in bias in 1991 is -20% . The estimated of growth rate over the time series using the average of the TVOD and RV series was unbiased at 2.0%. The estimates for each of the two decades were slightly positively biased. (b) The jump in bias in 1991 is $+20\%$. The estimated of growth rate over the time series using the average of the TVOD and RV series was unbiased at 2.0%. The estimates for each of the two decades were slightly negatively biased.

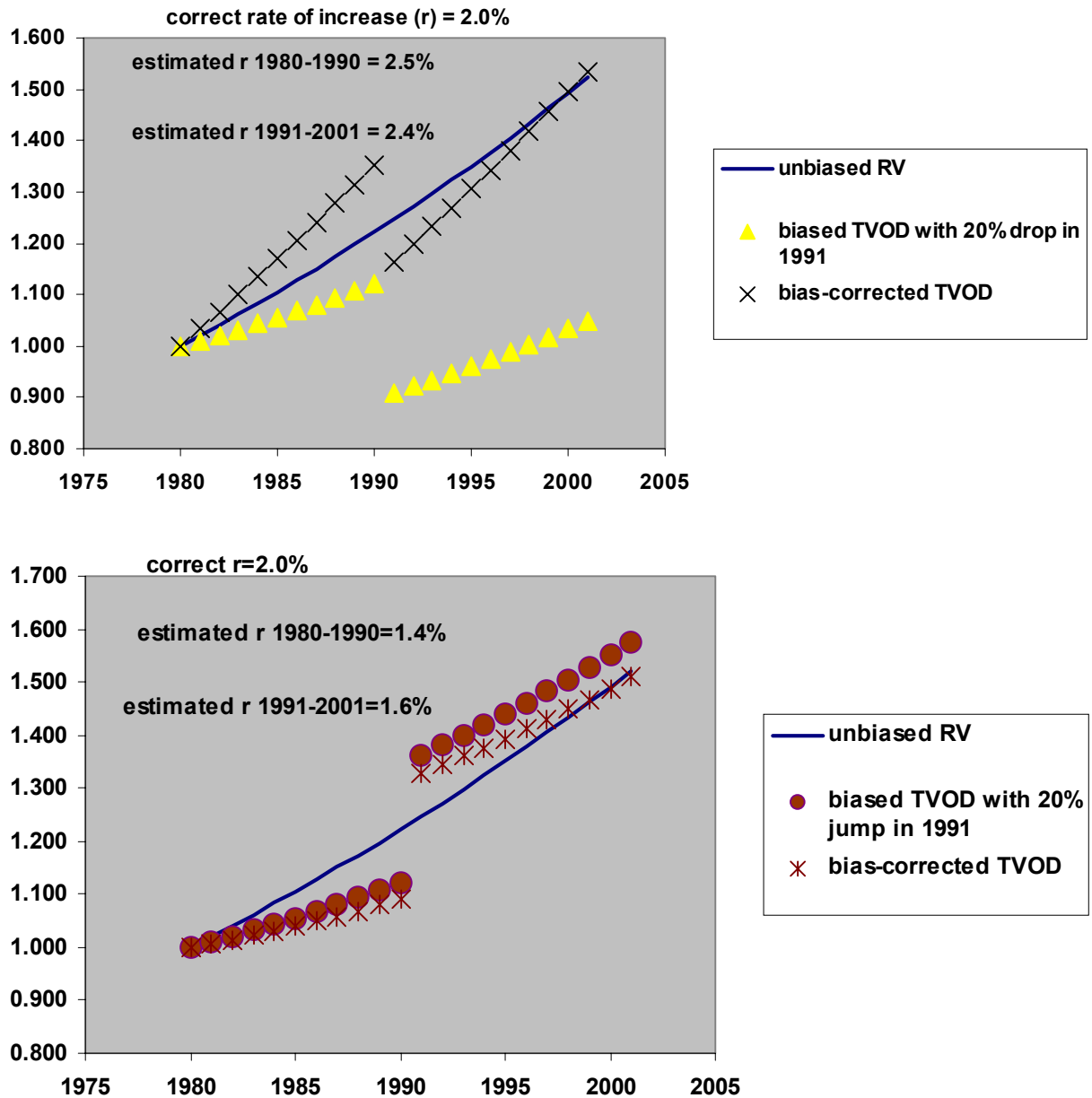
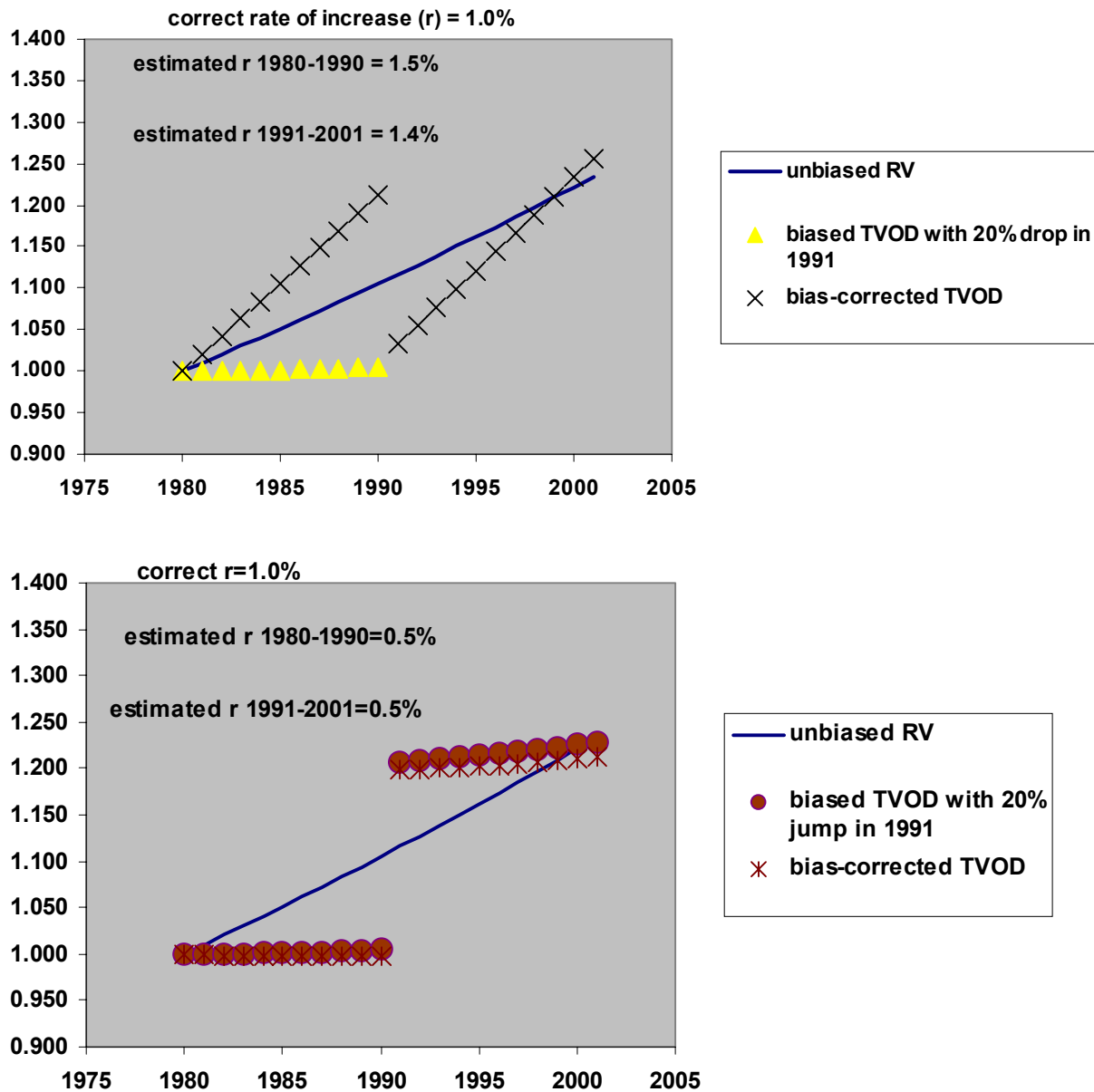


Figure 3. Deterministic trajectories of an unbiased RV time series together with a biased TVOD and bias-corrected TVOD time series. The true growth rate is 1.0%, the steady change in bias is -1% per year. (a) The drop in bias in 1991 is -20% . The estimated of growth rate over the time series using the average of the TVOD and RV series was unbiased at 1.0%. The estimates for each of the two decades were slightly positively biased. (b) The jump in bias in 1991 is $+20\%$. The estimated of growth rate over the time series using the average of the TVOD and RV series was unbiased at 1.0%. The estimates for each of the two decades were slightly negatively biased.



(2) None of the analyses of the TVOD data in Lennert-Cody et al. (2001) suggest abrupt changes in bias trends. The authors instead conclude that considering the variety of contributing sources of biases over the time series, the net changes in bias are expected to be gradual. The authors conclude by suggesting relatively small percentage gradual changes in biases per year over the time series of about -1% to -1.5%

at most. All of their empirical analyses supported only gradual changes in bias. The plausibility of the occurrence of abrupt changes in the TVOD time series was summarized p. 18 of my April Report:

A constant linear correction for the fishery dependent index should be considered for the following reasons:

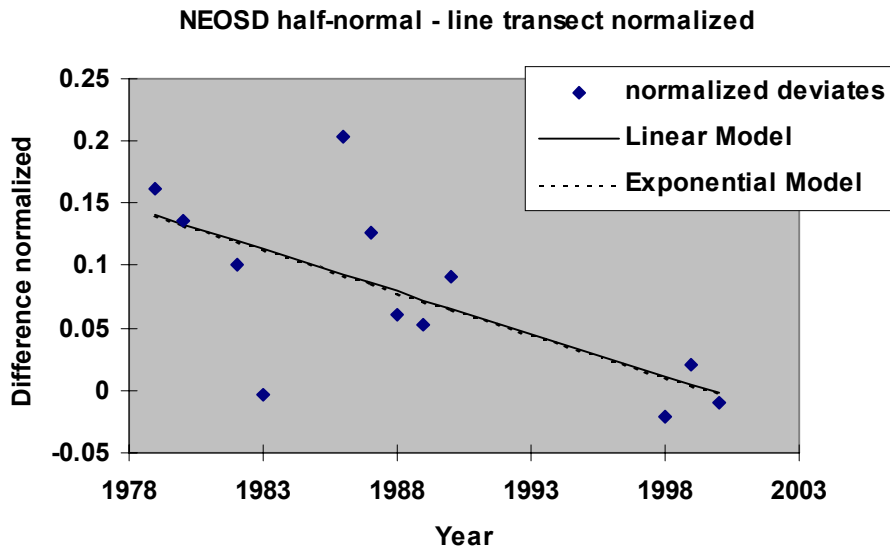
(1) **Conceptual support.** The thorough evaluations of the fishery dependent indices in Lennert-Cody et al (2001) indicated that the mechanisms most likely to cause temporal changes in bias in the fishery dependent (fd) indices were operating in a gradual manner over time, rather than at fixed and abrupt points in time. In other words, the causes of the biases were concluded to result from progressive changes in data quality and gradual fishery introduced-biases (please see the last two paragraphs on page 13 above for further details). All of the mechanisms identified to create trends in bias were found to create a negative bias that were expected to increase gradually in magnitude between 1979 and 2000. When considered to operate in combination, the combined effect of these various different mechanisms on the trend in bias should be that bias changes gradually not abruptly at certain points. The idea that *step changes* at either the imposition of regulatory measures that would affect reporting or at years of the major climatic regime shifts were *not* emphasized as being important in Lennert-Cody et al (2001). Moreover, if the latter were to occur, then the trend in the fishery independent indices would also be biased and need correction. Gradual changes in bias can be modeled by either a linear or exponential model.

(3) When compared to the RV time series, the TVOD series for ETPD and ESD do not exhibit any abrupt divergences in the late 1980s or early 1990s (see Fig. 5 below). If the RV time series are considered to be unbiased in their time trends, a gradual change in bias in the TVOD series over time would create an approximately linear or monotonic curvilinear pattern in the deviates between the TVOD and RV series. An abrupt change in bias in the TVOD series would be expected to create a distinctly non-linear, non-monotonic pattern in the deviates (see Figs. 1-3). When deviates between the TVOD and RV series are plotted over time, the deviates exhibit a linear relationship with time and do not show any abrupt changes at any point in the time series (Fig. 4). An exponential model was also fitted to the deviates between the TVOD and RV observations for NEOSD and ESD and the trends obtained were indistinguishable from the linear models (Fig. 4). When Bayes' factor was computed with use of my own importance sampling software to compare the linear model with the exponential model for both ETPD and ESD, Bayes' factor was 1.3 in favor of the linear model. Given that the exponential models obtained were indistinguishable from the linear models that were fitted and Bayes' factor favored, albeit only very slightly the linear models, it can be concluded that these results lend support for the use of a linear model of trend bias and make it unlikely that a single large abrupt change in bias in the TVOD time series has occurred. From p. 18 of my April 2002 Review:

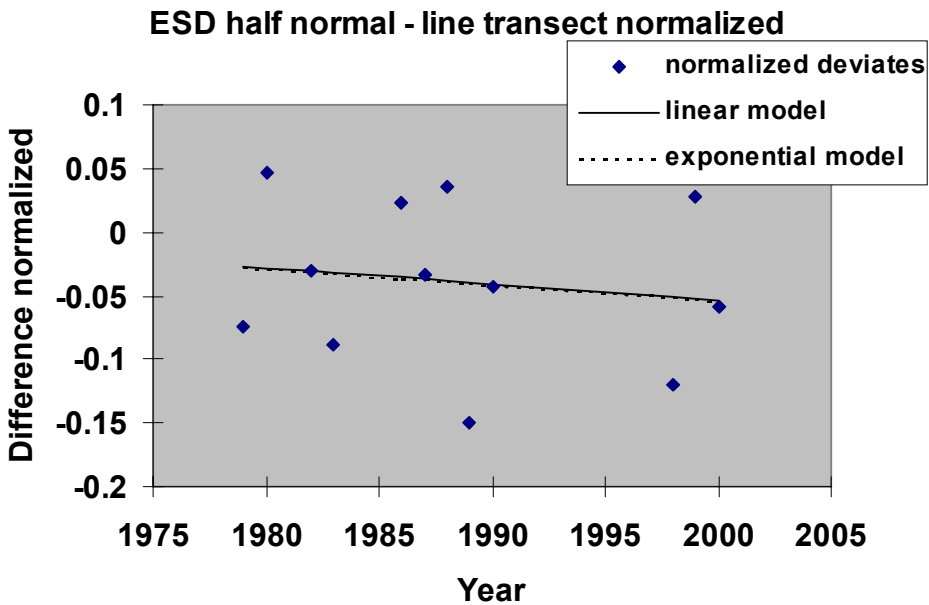
(2) **Empirical evidence and parsimony.** A useful empirical diagnostic to evaluate whether a linear model is appropriate is the plot of the differences between the fishery independent (fi) and fd indices against time as indicated above [Fig. 4]. In the ETPD assessment, the fi data are considered to be bias-free in their trends, and this notion is perfectly reasonable as it is in many other stock assessments. If this notion is not contested, then the above empirical comparisons between fi and fd indices for ETPD for years 1979-2000 [Fig. 4] strongly suggest that if there are temporal changes in bias in the fishery dependent (fd) indices, then these can be described adequately by a linear model over this period. In other words, on average, the absolute deviation between the true abundance and the fd increases in a linear fashion [Fig. 4] and the deviation has a negative trend. An exponential decline model could also be fitted to the data and a comparison with a linear model would require Bayes' factor for a probabilistic test of goodness of fit. However, the pattern in the differences [Fig. 4] appears to be linear and not exponential.

Figure 4. Plots of normalized differences between the fishery dependent and fishery independent indices. Best fitting linear and exponential models are also plotted. a. NEOSD half-normal – line transect. b. ESD half-normal – line transect (From Fig. 5, p. 16-17 of my April report).

a.

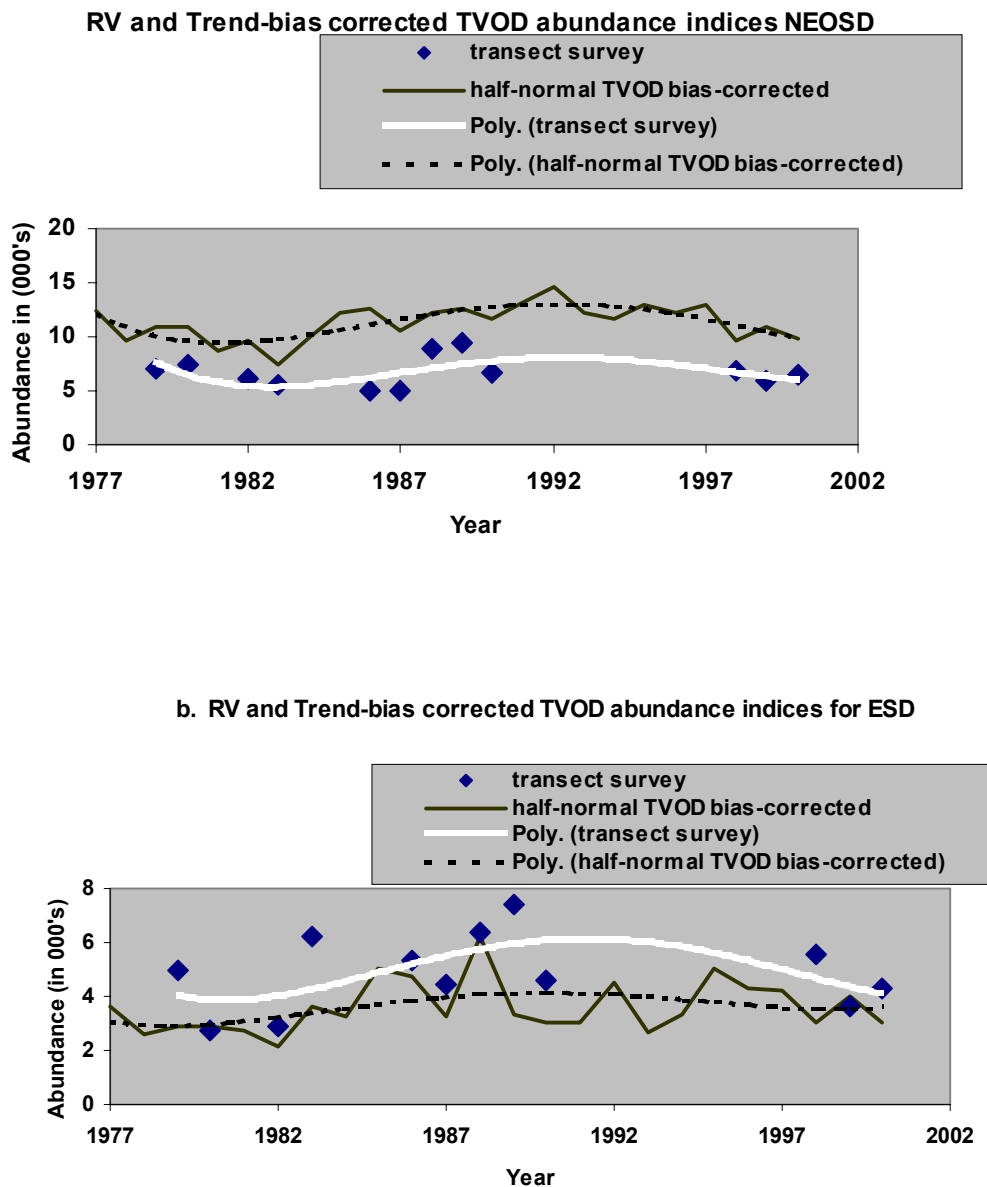


b.



(4) When the bias-corrected TVOD time series are superimposed over the RV time series, there is a very close match between the temporal patterns from the start of both time series until the end of each. In other words, the data in both time series appear to be tracking the same phenomena throughout the entire time series, allowing for random deviations due to random sampling error in each of the time series. The bias-corrected TVOD time series show no meaningful departures from the RV time series (Figures 4, 13 Wade July 2002, Fig. 5 below).

Fig. 5. RV and trend-bias corrected TVOD series and fitted polynomials for (a) NEOSD and (b) ESD (Lennert-Cody et al. 2001).



Taken together, these facts make it unlikely that (1) an abrupt change in bias in the TVOD time series could create an artificial estimated change in growth rate and thereby present a fundamental problem for the stock assessment, (2) a large abrupt change in bias has occurred over a short period in the TVOD time series and (3) a large abrupt change in bias in the TVOD time series some time between 1991 and 1998 has created an artificial, estimated change in growth rate in the late 1980s or early 1990s. A detailed set of reasons for why it is unlikely that the changes in bias in the TVOD time series are abrupt was provided in my April 2002 review. However, no rebuttal of these reasons has been provided in the response by the SWFSC. Instead, the draft appears to repeat the same speculative concerns voiced at the outset of the April 2002 review and lacks cognizance of the key features of the statistical methodology recommended to incorporate the TVOD data and the basic empirical evidence at hand that directly compares the RV and TVOD time series. It is surprising that no spreadsheet modeling or empirical comparisons of the RV and TVOD time series have been conducted to evaluate the validity of the speculative concerns expressed. When these were tried in this review, albeit simply, it was found that the concerns expressed were not born out, i.e., abrupt changes in bias did not create any meaningful changes in the expected average growth rate from the modeled TVOD and RV series.

In the telephone conference call on Friday August 16, a few of the NMFS scientists expressed a positive response to the April 2002 stock assessment review recommendations upon consideration of the various points made in an earlier draft of the current review. However, one of the NMFS scientists expressed some additional reservations. One was that the inclusion of a bias correction term for the TVOD series in the stock assessment model raised some philosophical questions about the validity of the statistical estimation procedure applied. It was questioned whether the likelihood function applied that assumes independence of each data point in each time series was still valid. This was presumably because one interpretation of the trend bias estimation suggested that the time trend in the TVOD series was effectively fitted to the RV time series. Upon further consideration, there appears to be no problem with the likelihood functions applied. By definition, a likelihood function gives the probability of obtaining a given observation or set of observations given, or conditioned on, a fixed set of values for the model's parameters. The data are treated as the random variable and the probability distribution for them can only be defined based on a fixed set of parameter values and a given model structure. Thus even when a bias function and trend bias parameter are included in the stock assessment model, the likelihood function for a TVOD observation should give the probability of obtaining this observation given a particular set of values for the estimated stock assessment model parameters. The likelihood of a given set of TVOD observations should be at its maximum when a set of values for model parameters is found that provides a best fit between model predicted and observed TVOD values. The same goes for the likelihood function of a RV observation. Thus, the likelihood functions are still conditioned on the values of parameters included in the stock assessment model and do not need to be reformulated when a bias function is incorporated.

Moreover, as in all instances when stock assessment models are fitted to cpue time series, a scale parameter, the constant of proportionality, must be estimated to scale the cpue time series to the model predicted abundance. The likelihood functions applied for the cpue data typically assume that each time series is independent, even when fishery independent abundance time series are also included in the stock assessment. The inclusion of an additional scale parameter that scales the trend in the cpue to the model predicted trend in abundance, although its estimation might be informed by a RV time series, does not change the independence of the likelihood functions. Not surprisingly, this conventional assumption of independence of likelihood functions for different cpue series has been applied in recent stock assessments in which trend bias functions for commercial cpue time series were explicitly modeled (Fournier et al. 1998; Parma 2001, 2002; Brooks et al. 2002; Porch 2002). Therefore, the conventional formulation of the likelihood function should hold even when TVOD series are also included in the stock assessment and trend bias is modeled in the TVOD.

It was also suggested by one of the NMFS scientists in the conference call that there was a strong correlation between the TVOD indices and the number of sets made, especially in the latter part of the time series, and that rather than tracking abundance, the TVOD might instead be tracking the number of sets. However, in the call Tim Gerrodette pointed out that in the most recent set of TVOD indices, there was a divergence between the number of sets and the index, thereby refuting this supposition.

Taking all of the evidence into account, there are in my opinion more compelling reasons to include the TVOD time series in the stock assessments than to leave them out. This is provided that statistical methodology is incorporated that allows the TVOD series to be used in a "data-filling" manner to fill in the large gaps in the RV data rather than as a time series that has equal status to the RV series in the estimation of population growth rate. This is also provided that diagnostics are used to ensure that the TVOD data fulfill no more than a data-filling role in the estimation. This data-filling role is supported firstly by the strong similarities in the plots against time of the bias-corrected TVOD and the RV time series; both time series for both ESD and NEOSD appear to be tracking the same underlying temporal pattern. Secondly, the model-projected abundances at least for the 2 slope exponential model correspond well to both the RV and TVOD series for both species. Thirdly, the exponential model's growth rate estimates are very similar with and without the TVOD series. Fourthly, the increase in the Bayes' factors from equivocal or very slightly in favor of, to slightly in favor of, a change in growth rate also suggests that the TVOD series take on no more than a data-filling role in the estimation.

The reluctance of scientists to apply the TVOD data in the assessment appears to arise from a concern that the choice of whether to incorporate the TVOD data implies an unacceptable bias-variance trade-off. On the one hand, leaving out the TVOD series and using only the RV time series is believed to give minimally biased estimates of growth rates and abundance trends. However, due to the low precision in the fishery independent RV observations and the very large gaps in the time series, especially between 1991 and 1997, the assessment produces only imprecise estimates of growth rates and is largely inconclusive for the purpose of comparing the plausibility of structurally different models. Thus the results obtained are deemed to be reasonably unbiased but suffer the problem of being imprecise and largely inconclusive with regards to several of the key questions addressed.

On the other hand, the inclusion of the TVOD time series can lead to more precise estimates and more powerful model discrimination because the TVOD observations are more precisely estimated and are available for each year in the time series from 1977-2000. Concerns have been expressed about the potential biases that could be introduced by the TVOD because they are fishery dependent indices and not fishery independent ones. A chief concern is that while more precise and conclusive results could be obtained, there is a risk that they will be unacceptably biased due to the manner in which the TVOD series were obtained and put together.

A statistical protocol however has been recommended and implemented that is specifically designed to eliminate trend-biases and yet retain the precision and model discrimination enhancements that could be obtained by the inclusion of the TVOD data in the stock assessment. True, not all of the precision and model discrimination enhancements are captured because the TVOD data are given a lower status in the estimation than the RV data and this weakens these enhancements. Also, the method cannot entirely eliminate bias but it should remove most of it, and do so even if large, abrupt changes in bias exist in the TVOD data (as demonstrated above). The potential for introduced bias can be evaluated simply by comparing the unadjusted TVOD time series with the RV series. Should any substantial non-linear or non-monotonic patterns be found in the deviates between the two series, there would remain a potential basis for bias to occur. But if the deviates are found to be linear over the time series (as they have been), it is unlikely that the TVOD series can introduce meaningful biases in the stock assessment. The potential losses in accuracy with the inclusion of TVOD are thus minor when compared to the important gains to be had in ability to discriminate between structurally different models (already shown for the exponential

model). Therefore, with this application of this statistical protocol to incorporate the TVOD series, the bias-variance trade-off is actually negligible. The protocol instead offers a basis to include the TVOD time series without introducing any meaningful biases and yet increase the precision and more importantly the model discrimination properties of the stock assessments.

It is thus recommended that all of the stock assessment modeling be done using both the TVOD series and the RV series and the recommended statistical methodology to incorporate the TVOD series. It is recommended that the results obtained from these revised analyses replace those using the RV series only because the RV series lead largely to inconclusive and equivocal results and the inclusion of the TVOD series in the manner recommended should not introduce bias but only enhance the statistical power of the tests of various hypotheses.

Detailed Comments on: Assessment of the population dynamics of the northeastern offshore spotted and the eastern spinner dolphin populations through 2002

Paul Wade. July 2002. Administrative Report LJ-02-13

Examined existing data for

- trends in population sizes, estimated average and maximum possible growth rates,
- compared current abundance levels to estimates of pre-exploitation abundance,
- tested for possible changes in growth rates and carrying capacity during the period since the onset of the fishery.
- examined the effects of including proxy measures of unobserved or unreported mortality from the tuna fishery
- population abundance from some model runs were projected 200 yrs into the future, given model parameters estimated from the data, to "estimate" how many years it might take for these depleted stocks to recover, i.e., return to optimum sustainable population levels (between 60% and 100% of pre-exploitation levels) **what specifically is the definition of OSY for ETPD?**
- three dolphin stocks classified as depleted under terms of the US Marine Mammal Protection Act:

NEO Spotted dolphin, eastern spinner dolphin, coastal spotted dolphin

p. 2. par. 2. line 5. "latter sub-species" instead of "sub-species"

- report addresses only two other stocks, which are the primary dolphins involved in the industry's capture of tunas in the ETP

p. 2. end "Following the methodology of Smith (1983), a revised assessment of the depletion level of eastern spinner dolphin was carried out after the depletion of abundance surveys from 1986-1990." This wording is confusing and could be reworded to start like: These methods followed that of and carried out a ...

p. 3. par. 5. Use of Bayes' Factor – actually Bayes' posterior for alternative models was recommended but similar idea.

p. 3. need to distinguish Bayes' Factor from Bayes' marginal posterior for a given model.

p. 4. analyses using the exponential and generalized logistic models were repeated incorporating TVOD estimates of relative abundance. – only the exponential model results are reported, what happened to the GLM results using TVOD? - why not apply also to age-structured models?

p. 4. par. 3. potential affect – potential effect

p. 4 McAllister (2002) CITED but not included in reference list.

p. 5. par. 2 from bottom. testing for possible additional mortality as a result of the fishery in addition to reported numbers. – need to explain why 50% and 100% values were chosen as multipliers for observed fishing mortalities. Reason for this analysis clear but justification of how it was done needs a bit more explanation.

- p. 5 bottom – 200 yr projection – need to justify why single and two r_{\max} cases were selected only for this. why not use the age-structured model for projections and what about use of Bayes' posteriors to evaluate relative plausibility of age-structured and non-age structured models?

p. 6. par. 3. (2) the fishing mortality was (????), please fill in the blank.

p. 6. bottom

$$M_y = M_0 \exp(H * E_y)$$

why this particular form and not a form like

$$M_y = M_0 + H * E_y ?$$

p. 7, middle: why is y_0 not the first year in which both series of indices have a data point?

p. 8, end of methods: what were the prior pdfs for each of the estimated parameters, e.g., a_{bias} ?

p. 9 top 2 par. This is an appropriate interpretation of Bayes' factor.

p. 9. par. 3. $CV_{\text{add}} = 0.33$ – please show equation for likelihood function of TVOD data, particularly the formulation of the variance in the LHF. also what is the prior for CV_{add} ?

"This value ensures that the TVOD estimates have the same weight in the results as do the research vessels estimates".

- In review it was recommended that the TVOD estimates be given no more weight than the research vessel estimates, not the same weight. It is not clear whether the priors for the CVs for the TVOD estimates is set up to ensure "the same weight" is given to them or the same or less weight, as recommended in the review. If it is intended that the same weight is given to them, then NMFS needs to justify this.

p. 9. par. 4. mention that posterior for case with TVOD data should be narrower. Should report the posterior CV in the estimates of r in the tables to facilitate the comparison of the estimates. 95% PIs are slightly narrower but PI's less precisely estimated than CVs

- interpretation of help of TVOD data in parameter estimation ok. It is correctly pointed out that if the fine-scale variability of the TVOD can provide a better fit to the model than is the case with the RV estimates, only then will the model results be enhanced over the case with only the RV data. The TVOD and RV time series when the model is fitted to them give similar time trends and temporal pattern and the temporal pattern in the data do not match well the temporal pattern indicated by the model which incorporates the bycatch by year to produce its temporal pattern in abundance. The model structure in this case does not allow the model to fit the data well because it does not account for the high point in both data series in the mid-80s and parameter estimation is not improved by also including the TVOD series. This fine resolution of the TVOD data, however, did help to improve parameter estimation and model discrimination when the 2-slope model was fitted to both the TVOD and RV data.

p. 9. bottom and first two par. of p. 10. The main finding here is that the addition of the TVOD series helped to improve model discrimination in favor of the two-slope model. The interpretation provided for this appears to be fine. The result was predicted in my initial review of the stock assessment. The temporal pattern in the RV data was fitted better by the two-slope model than the one-slope model but due to the sparseness of the data, the Bayes' factor computed using only the RV data was not conclusive. The temporal pattern in the TVOD data was similar to that in the RV data and the addition of the TVOD series, once it was trend-bias corrected also provided evidence more consistent with the two-line model. This increased the empirical basis and the Bayes' factor for the two-slope model.

p. 9. par. 2. the lower 95% probability interval (PI) for r_1 - r_2 is -0.02 . This means that at the lower bound r_2 was slightly larger than r_1 , i.e., that the rate of increase increased. Therefore, the last sentence of the paragraph is not technically correct. The decline in population growth rate in the latter period in the time series at the 0.975 probability level is by at least -0.02 not 0.02 as stated. It might be better to use a smaller cumulative percentile e.g., 90% so that the statement could read more intuitively with a positive decline (e.g., 0.02) rather than a negative decline (-0.02).

Why was the TVOD data used only with the exponential models? Conclusions using Bayes' factors could be substantially different for several of the model comparisons if TVOD data had been used also.

p. 9. last par. 1st sentence. "In this case, the TVOD data do lead to more precise results, and lead to more conclusive results that the population was increasing in the early period and declining in the later period." results are not conclusive. BF of only 5.11 does not provide conclusive findings

p. 10. 2nd par, 1st sent. "The model comparison gives a more conclusive result [than] did the model comparison using only research vessel estimates." This is an over-statement of findings.

same par. "... the analysis now leads to a conclusion that the data suggest the growth rate of the population did change in the [late] 1980s or early 1990s, but his results is not strong." Concluding that the data suggest change is perhaps an overstatement.

same par. There is an incorrect probability statement in the last sentence.

p.11. par. 3. "However, the comparison of these models' results to the model we ran with the actual reported fishery mortality result in low Bayes' factor, indicating that the base run of the generalized logistic model described above was favored over these models." The term favored should be qualified to slightly favored.

p. 13. top "and the effective "Rmax was lower during the latter half of the population trajectory (Table 16)." this Rmax was not given in Table 16.

p. 13. par. 1 "In other words, the fine-scale dynamics provided by the covariate do not match the available data better than the age-structured model that has survival constant across the time period". – True, but

this statement needs qualification – other model forms for the unobserved mortality rate model could potentially provide better fits to the data.

p. 16. bottom "comparison of these models' results to the model we ran with the actual reported fishery mortality resulted in low Bayes' factor, indicating that the base case run of the generalized logistic model described above was strongly favored over these models (Bayes factor of 2.43 for the 50% increase in fishery mortality scenario and 7.43 for the 100% increase scenario)." The statements are inconsistent with guidelines on the interpretation of Bayes' factors; this is over-stating the results.

p. 28, tab 17

- when TVOD data used, BF = 5.11 in favor of the 2 slope exponential model with the use also of the TVOD data; the BF was only 2.12 when only the RV data were used.

- This indicates that data are more powerful for model comparisons when TVOD data are used.

- p. 29. in TVOD runs, if CV_{add} is just an additive component of the variance, why is the CV_{add} for TVOD so large? (0.315, 0.301-0.388) (This seems precisely estimated). Please show the equations applied that use CV_{add} .

- What were the convergence diagnostics for SIR using the TVOD, and "M_y" model?

- p. 40. fig. 4d, e show that there is insufficient information in the data to estimate K2 parameter and break year. p. 49 same with fig. 15 d,e

- fig. 3, 4, show that trends in LT and TVOD observed series give consistent trends. For ESD and NEOSD, trends are consistent with modeled trends in abundance. Fig. 4 shows more consistency of 2 slope exponential to data than the 1 slope exponential to the data. This backs up Bayes' factors support for the 2 slope model. However, evidence should still be considered to be equivocal.

- the posterior for the year of change favors later years of change, after late 80s with peak 95-97. There appears to be bimodality – peak over 89 also – is this bimodality real or due to imprecision in importance sampling?

- NE OSD fig. 12. shows TVOD and LT RV observations are consistent in trend indication but that model trajectory and observations not consistent 1985-1995.

- fig. 13 both TVOD and LT indices show similar trends and are consistent with 2 slope model

- fig. 17, 18 – The trend from age structured model are not consistent with data 87-90 – ok.

p. 19. par. 3. "When the model is additionally fit to the TVOD estimates, the results do lead to the conclusion that the 2-slope model is better". Again, this is an overstatement of the results.

- p. 19 par. 3 the statement about the change in growth rate in early to mid 1990s is not consistent with posterior pdf in fig. 4 which shows peaks over early and late 90s. Also, it should be indicated that the median value for r_2 is -0.033 , to be consistent in reporting of results between NEOSD and ESD (latter report in conclusions a -0.04 rate of decrease, p.19, par.5).

p. 20. par. 2 "The standard model was estimated to have a probability 5 times greater for the covariate model, which is strong enough to be considered a positive result in favor of the standard model." Again, this is an overstatement of the results.

p. 20. par. 5. "When the exponential models were additionally fit to the TVOD estimates, the results did lead to the conclusion that the population growth rate had changed. " Again an overstatement of the results.

p. 21. par. 1 "That is, the modeling conducted here does not provide support for the possibility that unobserved or unreported mortality is occurring at those levels." Again this is an overstatement of the results.

p. 21. par. 3. "The standard age-structured model with constant survival provided a better fit to the data than did the age-structured model incorporating per capita sets per dolphins as a covariate for survival. ... which is strong enough to be considered a positive result in favor of the standard model" Again this is an overstatement of the results.

- p. 21 Malcolm Haddon, not Malcom Haddon

Detailed Comments on: Draft Report of the Scientific Research Program under the International dolphin conservation program act (August 2, 2002).

p. 8. The "confidence interval" CI for depletion is actually a probability interval (PI). The wide bounds in the PI could be legitimately narrowed if TVOD were used as they were with the exponential population dynamics models.

p. 9. par. 2. The following statements within the quote immediately below (particularly those in italics/bold and after) should be revised to account for the facts that a new statistical methodology was applied to remove potential trend-biases in the TVOD data and, upon inspection, there were no substantial differences in the fine-scale (e.g., within five-year) temporal patterns between the bias-corrected TVOD time series and the RV time series.

From p. 9. par. 2: "However, including the TVOD abundance indices indicated that there was a decrease in growth rates beginning around 1990. ***Whether this is an artifact of a changing bias over time within the TVOD time series is presently unknown***, but clearly this issues bears further scrutiny. Although the model runs using TVOD resulted in somewhat improved precision over those with only research vessel abundance data, we remain concerned about drawing conclusions from model runs using these estimates. In fact, significant problems are widely recognized in the TVOD series."

To the contrary, based on the new statistical methodology to incorporate the TVOD time series in the stock assessment and inspection of fine-scale temporal patterns in the bias-corrected TVOD series and RV series (and Excel Spreadsheet calculations of various scenarios for changes in bias in the TVOD data), it can be confidently concluded that the estimated changes in growth rates over time are not artifacts of a changing bias over time within the TVOD time series. This was explained in detail further above and the point is reiterated below.

The concerns raised in the quoted paragraph reflect issues raised in a recent paper that analyzed in detail the TVOD time series for ESD and NEOSD (Lennert-Cody et al. 2001). The analyses in Lennert-Cody et al. (2001) demonstrate empirically that there could be trend-biases in the TVOD time series, i.e., systematic deviations in the TVOD time series from the actual trends in abundance. The authors give indications of the potential sources and magnitudes of these trend-biases and warn that it would be ill advised to use these data as indices of abundance in stock assessment. As pointed out in my April 2002 review, this particular recommendation, while well-meaning, is misguided because it ignores a growing

body of pragmatic methodologies devised to bias-correct potential trend-biases in fishery-dependent indices of abundance so that they can be used in fisheries stock assessment (Wang and Die 1996; Robins et al. 1998; Bishop et al. 2000; Fournier et al. 1998; Parma 2001, 2002; Babcock and McAllister 2002; Brooks et al. 2002; Porch 2002). Many different stock assessments incorporate cpue data that are considered to still have time-trend bias in them and a number of these also incorporate mechanisms to account for the potential time-trend bias in the cpue data. A recently implemented mechanism is to model a random walk term within the stock assessment model that allows the deviate between the expected observed value of the cpue and true relative abundance to take on a random walk (Fournier et al. 1998; Parma 2001, 2002; Brooks et al. 2002; Porch 2002). Such general methodologies that model trend bias in cpue within the stock assessment (e.g., Babcock and McAllister 2002) are appropriate for instances such as this in which both fishery-dependent and fishery-independent data are available. There is often good reason to believe that fishery independent (RV) indices do not have a trend-bias, as has been assumed in this stock assessment. Given this, one simple and pragmatic approach to correcting the trend-bias in the fishery dependent data is to quantitatively compare the trends and temporal patterns in the fishery dependent and fishery independent data.

The basic points in the previous paragraph were made in my initial CIE review in April 2002. Analyses of the temporal patterns in the TVOD and RV time series for ESD and NEOSD in my April review showed that a linear model with bias approximated by a linear function of time from the start of the time series provided a good approximation of the empirical deviates between the RV and TVOD time series for both ESD and NEOSD. This functional form for potential bias in the TVOD time trend was recommended and modeled within the revised stock assessment (Wade July 2002). The potential biases in the TVOD trends were estimated assuming that the RV series gave an unbiased estimate of trends in abundance. This simple procedure essentially fitted the TVOD time series to the RV series (using a trend-adjustment parameter for the TVOD series) to eliminate any time trend discrepancies in the TVOD time series relative to those in the unbiased RV time series. Thus if the RV series can be considered unbiased in trend, then the TVOD data were bias-corrected for possible biases in temporal trends. *Thereby the modeled trends in the TVOD data cannot be different from those in the RV series.* When the model-adjusted, and bias corrected TVOD series and RV series were compared for their fine-scale (with-in five-year) temporal patterns, *the fine-scale temporal patterns were strikingly similar, allowing for the larger imprecision in the RV estimates* (see Figs. 4, 13 Wade, July 2002 and Fig. 5 above). Moreover, constraints were recommended and introduced into the stock assessment such that the assessment also did not permit the TVOD estimates to acquire more weight in the estimation than the RV estimates.

The methodology applied for using the TVOD thus did not allow them to take on the same weight as the RV series in estimating the central tendencies for trends in abundance or growth rates. The TVOD series instead provided more observations where there were gaps in the RV time series and improved the empirical precision and basis for model discrimination in the assessment. As mentioned above, strikingly similar fine-scale temporal patterns between the TVOD and RV data for both ESD and NEOSD can be seen after the TVOD data are corrected for potential bias in trend and bias in scale (assuming as in the assessment that there is no trend-bias or scale bias in the RV series) and the TVOD data are super-imposed on the RV data (Figs. 4, 13 Wade, July 2002, Fig. 5 above). As mentioned above, allowing for the random sampling error inherent in the different time series, the data points in both series indicate strikingly similar fine-scale temporal patterns between the TVOD and RV series throughout the entire time series where overlap occurs (1978-2000). This is found for both ESD and NEOSD. *This finding suggests that providing that a suitable trend-bias correction is applied to the TVOD data, the addition of the TVOD data to the stock assessment will not introduce biases to the assessment; instead, these data will only help to improve the otherwise weak empirical basis for the stock assessment* which otherwise fits various stock assessment models to a single time series of imprecise and relatively non-informative "one-way trip data" (Hilborn and Walters 1992), i.e., only twelve imprecise abundance data points in the latter half of the history of fishery interactions, well after the main depletion has occurred.

When my recommended methodological approach to incorporating the TVOD data was applied by Wade (July 2002), the basic improvements to the assessment that were predicted in my review were realized. First, the estimates of growth rates were not significantly changed but made more precise. The precision in parameter estimates improved from relatively little to considerably in the exponential models examined. The precision and posterior means were similar when TVOD were and were not included for the one slope exponential model, presumably because this model fitted neither series that well. In contrast, for the NEOSD, the precision in the estimate of the growth rate for the first part of the time series, r_1 , in the two slope exponential model improved dramatically. The 95% PI without the use of TVOD was $-0.066-0.071$ and this reduced to 0.011 to 0.077 with the use of TVOD (Wade July 2002, Table 29). The mean increased from 0.026 to 0.046 . Additionally, the statistical power in model discrimination was improved considerably due to the use of a larger number of bias-corrected abundance observations in the statistical evaluation.

The idea in the second sentence of p. 9 par. 2 that the estimated decrease in growth rates beginning around 1990 could have been due to a changing bias over time in the TVOD series is therefore not factually correct. The notion is not consistent with the methods of stock assessment analysis used to obtain this result, spreadsheet modeling carried out in this review and the empirical patterns in the time series data as seen in plots of the data and the results (Figs. 4,5 above and Figs. 4, 13 Wade, July 2002).

Firstly, this decrease in growth rates is suggested by the stock assessment analysis with the RV data only (Wade, July 2002). However, the RV data are too sparse and too imprecise for the null model of no change in growth rates to be rejected. When the TVOD data are included together with a trend-bias correction, the empirical basis for rejecting the null hypothesis of no change in growth rates increased. This is because, unlike the RV series, the TVOD time series has observations for each year between 1977 and 2000 and the TVOD observations are more precisely estimated. Because the modeled trend in the TVOD data is fitted within the stock assessment to the trend in the RV data (Wade July 2002, p. 9 par. 4) and the TVOD and RV time series show remarkably similar temporal patterns after a linear time-trend-bias correction, it can be concluded that the notions expressed in the text on p. 9. par. 2 are not plausible. In other words, the stock assessment methods applied and resulting patterns in the fits of the bias-corrected TVOD data to the RV data make it

(1) unlikely that the addition of the bias-corrected TVOD to the stock assessment will introduce any appreciable bias in estimated trends in population abundance and growth rates, and

(2) unlikely that the estimated decrease in growth rates beginning around 1990 could have been due to a changing bias over time in the TVOD series.

This is so unless after correcting for any bias in overall trend, the TVOD time series suggests different fine-scale temporal patterns in abundance. However, this is not the case (see Figs. 4, 13 Wade July 2002).

It is thus recommended that the text in p. 9. par. 2 be revised and corrected to take these findings into account.

Appendix on Stock Assessment

p. 78-79. Simple trend analysis. This is a very useful and important addition to the stock assessment and the statistical power analysis reported in it complements the trend analysis. It is recommended that the trend analysis also be conducted by incorporating the TVOD data with the recommended bias correction and variance formulation to evaluate the effects on statistical power of incorporating the bias-corrected TVOD series in the trend analysis.

All throughout the results section, the use of the term *confidence limits* is not technically correct and should be changed to probability interval because the results are Bayesian and not frequentist.

p. 81. "The per capita number of sets on dolphins per year gradually increased over time, with a sharp increase in the mid 1980s and then a small decline after 1990." The posterior median value for the maximum in the time series should be reported here.

p. 87. "the results of this simple trend analysis were not sufficiently informative about whether or not recovery was occurring". This conclusion is inconsistent with the statements within the trend analysis section. They were more definitive stating at the end of p. 78 and start of p. 79 that for example "it is highly unlikely that either the northeastern offshore spotted dolphin or the eastern spinner dolphin stock was growing at a rate of 3% per year or more during this period. The power analysis indicated about a 67% probability of detecting a 2% per year growth rate, suggesting less strongly that the growth rate did not reach this level." The more positive findings of the trend analysis section should be conveyed in this final summary section.

p. 88. par. 1 "including the TVOD abundance indices indicated that there was a change in growth rates for the stocks during this time. Whether or not this result is an artifact of a variable bias within the TVOD data series is unknown at this time, but clearly this issue bears further scrutiny". The interpretation in the second sentence is incorrect. Inspection of the deviations between the unadjusted TVOD series and the RV series we find that deviations are linear with time as assumed, and spreadsheet modeling of hypothetical abrupt changes in TVOD indices demonstrate that it is unlikely that the estimated changes in growth rates during the period 1979-2001 were an artifact of variable bias within the TVOD data. These findings support the use of TVOD as recommended in the April CIE review and demonstrate that biases from their use as recommended are unlikely.

Conclusions/Recommendations

The main limitation of the stock assessment is the paucity and imprecision of the RV data. Only twelve estimates are available over twenty-two years and there is large gap with no data between 1991 and 1997. Moreover, there is no contrast in the data; they represent abundances only after the major tuna fishery depletions have taken place and exemplify a simple one-way trip from a depleted state (Hilborn and Walters 1992). These limitations in the data severely limit the statistical inferences that can be made using stock assessment modeling. Relatively few parameters can be estimated in a given stock assessment model and there will be very limited ability to evaluate the plausibility of alternative stock assessment models, for example those with and without breaks in growth rate, those with and without breaks in carrying capacity, those with and without unobserved mortality as a function of dolphin sets per capita, and those with and without an additional fishery induced mortality rate factor in the latter part of the time series. These limitations were born out with the results of the stock assessment modeling. Estimates of growth rates in all of the models were relatively imprecise, though all suggested very low or zero or possibly declining population growth rates in the latter part of the time series. In terms of model discrimination, only low to moderate Bayes' factor values were obtained indicating that no strong results were obtained regarding the plausibility of alternative stock assessment models. Only inconclusive and equivocal results were found regarding the plausibility of the various stock assessment model alternatives considered. Due to the limitations in the variety of stock assessment model variants evaluated, results regarding choices between models structures were in all cases inconclusive.

Although the TVOD data were recommended in the April 2002 review to be included in the stock assessments, this was done only for the stock assessments using the exponential model. When these data were included, they did not lead to any meaningful differences in estimates of growth rates in the one

slope exponential model. However, in the two-slope exponential model the suggested drop in growth rate in the latter half of the time series when only the RV data were used became more pronounced statistically. The PIs for the earlier and later growth rates no longer overlapped. The use of the TVOD also led to more powerful tests of alternative model structures. When only the RV data were included, Bayes' factor results were largely equivocal between the single slope and two slope versions of the exponential model for both ESD and NEOSD. The addition of the TVOD data led to results moderately in favor of the two slope model, suggesting a decline in growth rates in the latter part of the time series. If the TVOD data were included in the other stock assessments as suggested in my April 2002 review, the plausibility of alternative model structures could be evaluated with a higher degree of empirical and methodological rigor.

The TVOD data were used only with the exponential models presumably because of concerns stated over the potential for the TVOD series to introduce bias in the stock assessment. These various concerns were evaluated in a review of the most relevant literature, spreadsheet modeling and inspection and statistical analysis of the TVOD and RV time series. The findings of my review indicate that it is unlikely that the use of the TVOD series in the manner recommended in the April 2002 stock assessment will introduce meaningful bias in the stock assessment. Instead, it is found that expected potential improvements in the precision of parameter estimates, especially in the more highly parameterized models evaluated, and the ability to statistically discriminate among alternative population dynamics models (particularly ones with and without changes in growth rates) provide compelling reasons for the TVOD series to be included in all of the stock assessment modeling, in the manner suggested in my April 2002 review.

Although Bayes' factors were computed as recommended in the April 2002 review, to evaluate the plausibility of alternative model structures, the values for Bayes' factors obtained were misinterpreted. Values between 2.12 and 7.43 were incorrectly taken to provide strong or conclusive support in favor of various model alternatives. Many inappropriate conclusions were made based on these intermediate values for Bayes' factors.

A relatively limited amount of modeling was conducted to explore the implications of, and plausibility of, fishery-induced unobserved mortality rates. Yet fairly strong conclusions were made stating for example, "that the stock assessment modeling does not support the possibility that there is substantial additional mortality as a simple function of the frequency of sets per year." (p. 26). These conclusions are overstatements of the findings and should be revised to correctly reflect the limitations of the analyses and more correctly correspond to correct interpretations of the weak to moderate Bayes' factors obtained.

Some Key Recommendations

- (1) It is recommended that the TVOD series be included in all of the stock assessment modeling in the manner suggested in my April 2002 review and considerations regarding potential changes in population processes be revised to take these new results into account.
- (2) The interpretations of Bayes' factors should be corrected to correspond to the guidelines for interpretation on p. 76 of the Draft Scientific Report, and taking into account the considerations provided in this review.
- (3) The conclusions in the Draft Scientific Report regarding the stock assessment modeling results should be revised where necessary to reflect the corrected interpretations of Bayes' factor results and the limitations in the analyses conducted.

(4) The conclusions in the Draft Scientific Report regarding the unobserved mortality rate covariate model should be revised to be less dismissive of the unobserved mortality rate hypothesis and take into account the positive and significant relationship found between mortality rate and dolphin sets per capita.

(5) It is recommended that additional modeling be carried out to further evaluate the implications and plausibility of alternative fishery induced mortality models. Models that incorporate chases rather than sets per year, incorporate estimates of mean number of animals set upon per year, model different functional forms for unobserved mortality such as additive models and models that model calf mortality only as a function of per capita number of sets, should also be considered.

(6) It is recommended that more detailed results be reported on the implications of the additional mortality rate models with either sets per year or chases per year incorporated. These additional results should include the modeled annual mortality rates and values for R_{\max} over the time series and plots of mortality rates and R_{\max} against sets per capita.

In summary, the empirical basis for the current stock assessment of ETPD is extremely weak and it is recommended that NMFS improve the empirical basis of the stock assessment by including in it the TVOD series. Due to the imprecision and sparseness of the RV time series, the estimates of population growth rates using only these data will remain highly imprecise. As suggested by the statistical power analysis in the Draft Scientific Report, it could only be concluded with confidence that the population growth rates of NEOSD and ESD are unlikely to be larger than about 4% over the last few decades. It can also be concluded with some confidence that the NEOSD and ESD dolphin stocks are still below their OSY levels.

This still leaves open a wide variety of plausible scenarios and this large certainty is not helpful for the Secretary of Commerce's important decision to be made in December of 2002 about whether dolphin sets made by the tuna purse-seine fishery are having an adverse impact on depleted ETPDs. From this stock assessment it is not possible to conclude whether the stock is currently declining, staying at the same level, or increasing as it might be expected to from the relatively small numbers of reported dolphin kills in the tuna purse seine fishery. It is currently not possible to make conclusions about the plausibility of the hypothesis that dolphin sets are causing unobserved mortality rates that are impeding population recovery. Due to the sparseness and imprecision in the RV observations, attempts to evaluate the plausibility of alternative hypotheses about potential changes in population growth rates as a result of changes in the fishery or other processes in the latter half of the time series will fail to provide anything but equivocal results as they have done in the most recent assessment. Disappointingly, no useful insights can be gathered from this stock assessment regarding potential changes in growth rates, carrying capacity, and unobserved mortality rates due to chase and encirclement.

In light of the empirical analyses in this review and methods available to deal with trend bias in commercial cpue data, NMFS' recent position to not include the TVOD time series in the stock assessment of ESD and NEOSD is scientifically indefensible. This hesitation is founded on the findings from recent studies (e.g., Lennert-Cody et al. 2001) that suggest that there is trend bias in the TVOD time series. However, stock assessment methodologies have recently been developed to correct for trend bias in commercial cpue series (Fournier et al 1998; Parma 2001, 2002; Babcock and McAllister 2002; Brooks et al. 2002; Porch 2002) and it is recommended that the general approach of correcting for trend bias in commercial cpue data should be considered for this stock assessment. Empirical analyses in this review demonstrate that the trend bias is linear and once a linear bias correction is applied, the TVOD and RV time series very closely follow the same temporal patterns. Therefore, a linear trend bias correction factor is appropriate for use with the TVOD time series.

NMFS' recent position not to include the TVOD series in the stock assessment is impeding the scientific process and the ability to make scientific inferences about the current state of the stock and potential recent changes in population dynamics. Due to the considerable limitations imposed on statistical inference from the sparse and imprecise RV series, it is recommended that NMFS put aside its speculative hesitations about the use of the TVOD series and incorporate them in the stock assessment using the bias correction method suggested in my April 2002 review. The analyses conducted in this review indicate that it is unlikely that incorporating the TVOD series in the stock assessment in the manner suggested will introduce any meaningful biases in the stock assessment, providing that the RV data are unbiased. Instead, the incorporation of the TVOD series in the stock assessment will improve its currently very weak empirical basis and should help to improve the statistical power of evaluations of the plausibility of alternative hypotheses regarding the potential impacts of the tuna purse seine fishery on the population growth rates of ETPD. However, due to the limited role given to the TVOD time series, it is unlikely that they will provide any more than moderate and suggestive evidence regarding the plausibility of alternative hypotheses about potential recent changes in population growth rates. Nonetheless, the potential additional insights into ETPD population dynamics that might be provided by addition of the TVOD series to the stock assessment are worthy of further exploration and could potentially further assist the Secretary of Commerce in his upcoming decision regarding the potential impact of tuna purse seine dolphin sets on ETPD.

References

- Archer, F., Gerrodette, T. Jackson, A. 2002. Preliminary estimates of the annual number of sets, number of dolphins chased, and number of dolphins captured by stock in the tuna purse-seine fishery in the eastern tropical Pacific, 1971-2000. National Marine Fisheries Service Administrative Report, La Jolla, LJ-02-10.
- Babcock, E.A. and McAllister, M.K. 2002. Bayesian methods for accounting for data contradictions in stock assessment of Atlantic white marlin (*Tetrapterus albidus*). International Commission for the Conservation of Atlantic Tunas, Madrid. Document SCRS/02/067.
- Bishop, J., D.J. Die and Y. Wang. 2000. Using generalised estimating equations to assess increases in fishing power caused by changes in vessel technology. Australia New Zealand Journal of Statistics, 42: 901-919.
- Brooks, E., Porch, C., and Cortes, E., C. 2002. An age-structured production model (ASPM) for application to large coastal sharks. 2002 Shark Evaluation Workshop. National Marine Fisheries Service, Panama City. Document SB-02-5.
- Fournier, D.A., Hampton, J. and Sibert, J.R. 1998. MULTIFAN-CL: a length-based, age-structured model for fisheries stock assessment, with application to South Pacific albacore, *Thunnus, alalunga*. Can. J. Fish. Aquat. Sci. 55:2105-2116.
- Hilborn, R. and Walters, C.J. 1992. Quantitative fisheries stock assessment: choice, dynamics, and uncertainty. Chapman and Hall, New York, N.Y. 570 p.
- Kass, R. E. and A. E. Raftery. 1995. Bayes factors. J. Am. Stat. Assoc. 90: 773-795.
- Lennert-Cody, C.E., Buckland, S.T. and Marques, F.F.C. 2001. Trends in dolphin abundance estimated from fisheries data: a cautionary note. J. Cetacean Res. Manage. 3(3):305-319.

- McAllister, M.K. and Kirchner, C. 2002. Accounting for structural uncertainty to facilitate precautionary fishery management: illustration with Namibian orange roughy. In "Targets, Thresholds, and the Burden of Proof in Fisheries Management" Mangel, M. ed. Bull. of Mar. Sci., 70(2). 499-540.
- Parma, A. 2001. Bayesian approaches to the analysis of uncertainty in the stock assessment of Pacific Halibut. In J. M. Berkson, L.L. Kline and D.J. Orth (eds.) Incorporating uncertainty into fisheries models. American Fisheries Society Symposium, Bethesda, MD. 24: 111-132.
- Parma, A. 2002. In search of robust harvest rules for pacific halibut in the face of uncertain assessments and decadal changes in productivity, In "Targets, Thresholds, and the Burden of Proof in Fisheries Management" Mangel, M. ed. Bull. of Mar. Sci., 70(2). 423-453.
- Porch, C.E. 2002. A preliminary assessment of Atlantic white marlin (*Tetrapturus albidus*) using a state-space implementation of an age-structured production model. International Commission for the Conservation of Atlantic Tunas, Madrid. Document SCRS/02/068.
- Robins C.M., Y. Wang and D. J. Die. 1998. The impact of global positioning systems and plotters on fishing power in the Northern Prawn Fishery, Australia. *Can. J. Fish. Aquat. Sci.* 55: 1645-1651.
- SWFSC. 2002. Draft Report of the overall IDCPA research program and results (July 2002).
- Wade, P. 2002. Assessment of the population dynamics of the northeastern offshore spotted and the eastern spinner dolphin populations through 2002. National Marine Fisheries Service Administrative Report, La Jolla, LJ-02-13.
- Wang Y. and D.J. Die. 1996. Stock-recruitment relationships of the tiger prawns (*Penaeus esculentus* and *Penaeus semisulcatus*) in the Australian Northern Prawn Fishery. *Marine and Freshwater Research*, 47: 87-95.

Appendix 1: Statement of Work

Consulting Agreement Between The University of Miami and Dr. Murdoch McAllister

Background

The tuna purse seine fishery has used the association between tuna and dolphins to fish in the eastern tropical Pacific Ocean (ETP) for over five decades. Three stocks of dolphins were depleted by high historical levels of dolphin mortality in tuna purse seine nets, with an estimated 4.9 million dolphins killed during the fourteen-year period 1959-1972. After passage of the Marine Mammal Protection Act (MMPA) in 1972 and the increased use of fishing equipment and procedures designed to prevent dolphin deaths, mortality decreased during the late 1970s, 1980s, and 1990s to levels that are generally considered biologically insignificant.

While changes in the fishery have dramatically reduced the observed mortality of dolphins, the MMPA, as amended by the International Dolphin Conservation Program Act (IDCPA), requires that the National Marine Fisheries Service (NMFS) conduct research consisting of three years of population abundance surveys and stress studies to form the basis of a determination by the Secretary of Commerce regarding whether the “intentional deployment on, or encirclement of, dolphins by purse-seine nets is having a significant adverse impact on any depleted dolphin stock”. The Secretary must make a final finding in this regard by December 31, 2002. It should be noted that this issue is controversial and particularly relevant to persons involved with NMFS, the US and non-US tuna industry, and environmental groups.

The topic of this review is the IDCPA Science Report that will be presented to the Secretary of Commerce, along with information obtained under the IDCP, and other relevant information to form the basis of the Secretary’s final finding. The IDCPA Science Report is comprised of the results of all research activities required under section 304(a) of the MMPA, as amended by the IDCPA. Each major component of this report has been separately considered in a series of independent peer reviews conducted by the Center for Independent Experts (CIE). These consist of: the Abundance Review (October 15-17, 2001) the Stress Review (February 4-6, 2002), the Ecosystem Review (March 6-8, 2002), and the Assessment Model Review (April 3-5, 2002).

Abundance Review

The topic of this review was the abundance of several species of tropical pelagic dolphins that associate with tuna and are killed in the ETP purse seine tuna fishery. Estimates of dolphin abundance based on cruises carried out in 1998-2000 form a central part of these studies. The main task of the consultant was to review the methods used to estimate abundance from line-transect data, including covariate detection models. The fact that these dolphins occur in a wide range of school sizes presents unique problems for the estimation of expected group size, so considerable effort has been devoted to this analysis. Documents supplied to the reviewers included draft manuscripts describing the covariate analysis, simulations to test the performance of several estimators, calibration of school size estimates, and assignment of partially identified sightings. Background papers included previous relevant publications and reports. The raw data and software used in the analysis were also made available.

Stress Review

The stress studies mandated in the IDCPA include: 1) a review of relevant stress-related research and a three-year series of necropsy samples from dolphins obtained by commercial vessels; 2) a one-year review of relevant historical demographic and biological data related to the dolphins and dolphin stocks; and 3) an experiment involving the repeated chasing and capturing of dolphins by means of intentional encirclement. This review included a suite of studies subsumed under this general topic, and a brief description of these studies follows.

The necropsy program analyzed samples from about 50 dolphins killed incidentally during fishing operations. Historical biological samples and data were analyzed to investigate stress-activated-proteins (SAPs) in the skin in dolphins killed in the fishery and live-sampled via biopsy. Historical data were also examined to assess separation of cows and calves during fishing operations. Chase Encirclement Stress Studies were conducted during a two-month research cruise aboard the NOAA ship McArthur in the ETP. During this project, the team worked in cooperation with a chartered tuna purse seine vessel to study potential effects of chase and encirclement on dolphins involved in tuna purse seine operations. Dolphins groups were found to be much more dynamic than previously recognized, making it extremely difficult to recapture groups of dolphins over the course of several days to weeks, as planned.

In the end, nine different dolphins were tracked for 1-5 days during the course of the study, including two animals outfitted with a thermal tag that recorded heat flux, temperature, and dive data. Individual radio-tagged dolphins and 1-4 associated roto-tagged dolphins were recaptured on several occasions spanning shorter periods of 1-3 days. Six satellite tags were deployed to record movement and dive data on dolphins that were not recaptured. Biological data and samples were collected from as many captured dolphins as possible, and include: 70 blood samples, of which 18 were from repeat captures of marked individuals; 283 skin samples, of which 17 were from previously captured and sampled animals; 449 analyzable thermal images; 52 core temperatures; and 95hrs of heat flux data. Females with calves were noted on several recapture occasions, and one known calf was skin sampled during an initial and subsequent capture.

Ecosystem Review

To complement the three-year abundance studies, population assessments were made for the following years: 1986, 1987, 1988, 1989, 1990, 1998, 1999, and 2000 with a primary goal being to determine if populations that were historically reduced in size are increasing over time. Should the assessments indicate no increase (lack of recovery), three broad categories of factors could be the cause: a) effects from the fishery; b) effects from the ecosystem; c) an interaction between the proceeding two factors. This need to attribute causality for a potential lack of recovery serves as the primary justification for ecosystem studies. By investigating the physical and biological variability of the ecosystem of which the dolphin stocks are a part, we establish a context, which can be used to better interpret trends in dolphin abundance. A lack of recovery that is not mirrored by some other change in the ecosystem would largely eliminate an ecosystem hypothesis, leaving fishery effects as the most likely cause.

This review included a suite of studies subsumed under the general topic of ecosystem research in the ETP. The basic approach was to compare ecosystem parameters over time with a primary goal being to look for indications of a potential ecosystem shift. The power of these ecosystem studies increased with the number of environmental variables, taxa, and trophic levels included, and with the time period spanned (although most ecosystem data available for these investigations were collected concurrently with dolphin assessment data aboard NOAA research vessels and are restricted to the late 1980s and late 1990s).

The general components of the ecosystem research included: 1) physical and biological oceanography: sea surface temperature, thermocline characteristics, phytoplankton and zooplankton distribution and relative abundance; 2) larval fishes: distribution and relative abundance; 3) flying fishes: distribution, relative abundance, and habitat relationships; 4) seabirds: distribution, absolute abundance, and habitat relationships; and 5) cetaceans: distribution, absolute abundance, and habitat relationships.

Assessment Model Review

As indicated above, NMFS was charged with essentially determining whether or not the depleted dolphin stocks are recovering, and if so, at what rate and at what level of certainty. The topic of this review was the overall framework that will be to estimate the growth rate of two dolphin populations of interest, the northeastern offshore spotted dolphin and the eastern spinner dolphin, using growth rates estimated by fitting a population model to the three-year and other available estimates of abundance. For this review, estimates from research vessel surveys using line transect methods are available for three periods: 1979-83 (four estimates), 1986-90 (five estimates), and 1998-2000 (three estimates), for a total of twelve estimates over twenty-one years. Reviewers were also asked to evaluate the inclusion or exclusion of a set of fishery-dependent indices of abundance, resulting from data collected by tuna vessel observers. Two types of population growth rate will be estimated: (1) exponential rate of change from 1979-2000 and (2) intrinsic rate of increase under the assumption of a density-dependent model where pre-exploitation population size in 1958 is considered carrying capacity. Both an aggregated population model and an age-structured model will be used. Bayesian statistics, using a numerical integration method, were used to estimate a probability distribution for the population growth rate.

Specific Reviewer Responsibilities

For the final IDCPA Science Program Review, expertise is needed to review all components of the research described above, specifically with respect to NMFS' incorporation of comments previously received from the topical reviews also described above. Reviewers will be provided with the draft IDCPA Science Report, as well as comments received as a result of the CIE reviews and explanations of how/why such comments were or were not incorporated into the report.

The reviewer's duties shall not exceed a maximum total of 11 days, including:

- 2-3 days to read the draft IDCPA Science Report (to be provided to the reviewers by no later than August 2, 2002);
- 2-3 days to produce a written report of the reviewer's comments and recommendations on the draft report;
- 1-2 days to discuss via telephone, on August 15-16, 2002, with relevant NMFS staff from the NMFS La Jolla Laboratory, the incorporation of comments and any related questions; and
- 2-3 days to revise the written report based on those discussions.

It is expected that each reviewer will have participated in the earlier CIE reviews of IDCPA research described above and will not require general presentations of research results, but will focus on addressing comments and recommendations included in the reviewers' reports in his/her topic area. Reviewers should particularly consider whether the responses to the original review comments are sufficient and acceptable, in a manner similar to the role filled by a journal editor when considering manuscripts revised in response to referees' comments.

Each reviewer's report shall reflect the reviewer's area of expertise; therefore, no consensus opinion (or report) will be required. Specific tasks and timings are itemized below:

1. Read and become familiar with the draft IDCPA Science Report provided in advance;
2. No later than August 13, 2002, submit a written report of findings, analysis, and conclusion in the individual reviewer's topic area to NMFS;
3. Discuss relevant documents with scientists from the NMFS La Jolla Laboratory via telephone on August 15-16, 2002, to facilitate proper incorporation of reviewers' comments;
4. No later than August 23, 2002, submit a revised written report of findings, analysis, and conclusions based on discussions held with relevant NMFS staff from the NMFS La Jolla Laboratory. The written report¹ (see Annex I) should be addressed to the "University of Miami Independent System for Peer Review," and sent to Dr. David Die, via email to ddie@rsmas.miami.edu, and to Mr. Manoj Shrivani, via email to mshrivani@rsmas.miami.edu.

¹ The written report will undergo an internal CIE review before it is considered final. After completion, the CIE will create a PDF version of the written report that will be submitted to NMFS and the consultant.

Appendix 2: Bibliography of Materials Provided by the Center for Independent Experts

- Haddon, M. 2002. Report on the “Assessment of the Population Dynamics of the northeast offshore spotted and the eastern spinner dolphin populations through 2001”. CIE. Miami.
- Medley, P. 2001. Review of Transect Sampling Methods to Obtain Population Size Estimates for Northeastern Offshore Spotted and Eastern Spinner Dolphins. CIE. Miami.
- Mohn, R. 2001. Review of methods to estimate dolphin abundance from line transect surveys and presentation at Southwest Fisheries Science Center (SWFC), La Jolla, California 15-17 Oct. 2001. CIE. Miami.
- SWFSC. 2002. Draft Report of the overall IDCPA research program and results (July 2002).

2. Abundance estimates for depleted dolphin stocks:

- Brandon, J., T. Gerrodette, W. Perryman and K. Cramer. 2002. Responsive movements and $g(0)$ for target species of research vessel surveys in the eastern tropical Pacific Ocean. Administrative Report No. LJ-02-02, NMFS, Southwest Fisheries Science Center, 8604 La Jolla Shores Drive, La Jolla, CA 92037.
- Forcada, J. 2002. Multivariate methods for size dependent detection in conventional line transect sampling. Administrative Report No. LJ-02-07, NMFS, Southwest Fisheries Science Center, 8604 La Jolla Shores Drive, La Jolla, CA 92037.
- Gerrodette, T and J. Forcada. 2002. Estimates of abundance of northeastern offshore spotted, coastal spotted, and eastern spinner dolphins in the eastern tropical Pacific Ocean. Administrative Report No. LJ-02-06, NMFS, Southwest Fisheries Science Center, 8604 La Jolla Shores Drive, La Jolla, CA 92037.
- Gerrodette, T., W. Perryman and J. Barlow. 2002. Calibrating group size estimates of dolphins in the eastern tropical Pacific Ocean. Administrative Report No. LJ-02-08, NMFS, Southwest Fisheries Science Center, 8604 La Jolla Shores Drive, La Jolla, CA 92037.
- Kinzey, D., T. Gerrodette and D. Fink. 2002. Accuracy and precision of perpendicular distance measurements in shipboard line-transect sighting surveys. Administrative Report No. LJ-02-09, NMFS, Southwest Fisheries Science Center, 8604 La Jolla Shores Drive, La Jolla, CA 92037.

3. Ecosystem studies:

- Ballance, L. T., P. C. Fiedler, T. Gerrodette, R. L. Pitman and S. B. Reilly. 2002. An overview of eastern tropical Pacific ecosystems studies within the context of International Dolphin Conservation Program Act research. Administrative Report No. LJ-02-14, NMFS, Southwest Fisheries Science Center, 8604 La Jolla Shores Drive, La Jolla, CA 92037.
- Ballance, L. T., R. L. Pitman, L. B. Spear and P. C. Fiedler. 2002. Investigations into temporal patterns in distribution, abundance and habitat relationships within seabird communities of the eastern tropical Pacific. Administrative Report No. LJ-02-17, NMFS, Southwest Fisheries Science Center, 8604 La Jolla Shores Drive, La Jolla, CA 92037.
- Fiedler, P. C. 2002. Environmental change in the eastern tropical Pacific Ocean: Review of ENSO and decadal variability. Administrative Report No. LJ-02-16, NMFS, Southwest Fisheries Science Center, 8604 La Jolla Shores Drive, La Jolla, CA 92037.

- Fielder, P. C. and V. A. Philbrick. 2002. Environmental change in the eastern tropical Pacific Ocean: Observations in 1986-1990 and 1998-2000. Administrative Report No. LJ-02-15, NMFS, Southwest Fisheries Science Center, 8604 La Jolla Shores Drive, La Jolla, CA 92037.
- Gerrodette, T. and J. Forcada. 2002. Estimates of abundance of striped and common dolphins, and pilot, sperm and Bryde's whales in the eastern tropical Pacific Ocean. Administrative Report No. LJ-02-20, NMFS, Southwest Fisheries Science Center, 8604 La Jolla Shores Drive, La Jolla, CA 92037.
- Moser, H. G., P. E. Smith, R. L. Charter, D. A. Ambrose, W. Watson, S. R. Charter and E. M. Sandknop. 2002. Preliminary report on ichthyoplankton collected in manta (surface) net tows on marine mammal surveys in the eastern tropical Pacific: 1987–2000. Administrative Report No. LJ-02-18, NMFS, Southwest Fisheries Science Center, 8604 La Jolla Shores Drive, La Jolla, CA 92037.
- Pitman, R. L., L. T. Ballance and P. C. Fiedler. 2002. Temporal patterns in distribution and habitat associations of prey fishes and squids. Administrative Report No. LJ-02-19, NMFS, Southwest Fisheries Science Center, 8604 La Jolla Shores Drive, La Jolla, CA 92037.
- Reilly, S. B., L. T. Ballance, P. C. Fiedler, T. Gerrodette, R. L. Pitman, H. G. Moser, L. B. Spear and J. M. Borberg. 2002. Recent information to aid in detecting regime shifts in the eastern tropical Pacific Ocean. Administrative Report No. LJ-02-22, NMFS, Southwest Fisheries Science Center, 8604 La Jolla Shores Drive, La Jolla, CA 92037.
- Reilly, S. B., P. C. Fiedler, T. Gerrodette, L. T. Ballance, R. L. Pitman, J. M. Borberg and R. L. Holland. 2002. Eastern tropical Pacific dolphin habitats - Interannual variability 1986-2000. Administrative Report No. LJ-02-21, NMFS, Southwest Fisheries Science Center, 8604 La Jolla Shores Drive, La Jolla, CA 92037.

4. *Stress and other possible fishery effects:*

- Chivers, S. J., and M. D. Scott. 2002. Tagging and tracking of *Stenella* spp. during the 2001 Chase Encirclement Stress Studies Cruise. Administrative Report No. LJ-02-33, NMFS, Southwest Fisheries Science Center, 8604 La Jolla Shores Drive, La Jolla, CA 92037.
- Cowan, D. F., and B. E. Curry. 2002. Histopathological assessment of dolphins necropsied onboard vessels in the eastern tropical Pacific tuna fishery. Administrative Report No. LJ-02-24C, NMFS, Southwest Fisheries Science Center, 8604 La Jolla Shores Drive, La Jolla, CA 92037.
- Cramer, K., and W. L. Perryman. 2002. Estimation of reproductive and demographic parameters of the eastern spinner dolphin (*Stenella longirostris orientalis*) using aerial photography. Administrative Report No. LJ-02-31, NMFS, Southwest Fisheries Science Center, 8604 La Jolla Shores Drive, La Jolla, CA 92037.
- Dizon, A., A. Allen, N. Kellar, and Sárka Southern. 2002. Stress in spotted dolphins (*Stenella attenuata*) associated with purse-seine tuna fishing in the eastern tropical Pacific. Administrative Report No. LJ-02-26, NMFS, Southwest Fisheries Science Center, 8604 La Jolla Shores Drive, La Jolla, CA 92037.
- Edwards, E. F. 2002. Behavioral contributions to separation and subsequent mortality of dolphin calves chased by tuna purse-seiners in the eastern tropical Pacific Ocean. Administrative Report No. LJ-02-28, NMFS, Southwest Fisheries Science Center, 8604 La Jolla Shores Drive, La Jolla, CA 92037.
- Edwards, E. F. 2002. Energetics consequences of chase by tuna purse-seiners for spotted dolphins (*Stenella attenuata*) in the eastern tropical Pacific Ocean. Administrative Report No. LJ-02-29, NMFS, Southwest Fisheries Science Center, 8604 La Jolla Shores Drive, La Jolla, CA 92037.

- Forney, K. A., D. J. St. Aubin, and S. J. Chivers. 2002. Chase Encirclement Stress Studies on dolphins involved in eastern tropical Pacific Ocean purse seine operations during 2001. Administrative Report No. LJ-02-32, NMFS, Southwest Fisheries Science Center, 8604 La Jolla Shores Drive, La Jolla, CA 92037.
- Mesnick, S. L., F. I. Archer, A. C. Allen, and A. E. Dizon. 2002. Evasive behavior of eastern tropical Pacific dolphins relative to effort by the tuna purse-seine fishery. Administrative Report No. LJ-02-30, NMFS, Southwest Fisheries Science Center, 8604 La Jolla Shores Drive, La Jolla, CA 92037.
- Pabst, D. A., W. A. McLellan, E. M. Meagher, A. J. Westgate. 2002. Measuring temperatures and heat flux from dolphins in the eastern tropical Pacific: Is thermal stress associated with chase and capture in the ETP-tuna purse seine fishery? Administrative Report No. LJ-02-34C, NMFS, Southwest Fisheries Science Center, 8604 La Jolla Shores Drive, La Jolla, CA 92037.
- Romano, T., K. Abella, D. Cowan, and B. Curry. 2002. Investigation of the morphology and autonomic innervation of the lymphoid organs in the pantropical spotted, spinner, and common dolphins (*Stenella attenuata*, *Stenella longirostris* and *Delphinus delphis*) incidentally entangled and drowned in the tuna purse-seine fishery in the eastern tropical Pacific. Administrative Report No. LJ-02-25C, NMFS, Southwest Fisheries Science Center, 8604 La Jolla Shores Drive, La Jolla, CA 92037.
- Romano, T., M. Keogh, and K. Danil. 2002. Investigation of the effects of repeated chase and encirclement on the immune system of spotted dolphins (*Stenella attenuata*) in the eastern tropical Pacific. Administrative Report No. LJ-02-35C, NMFS, Southwest Fisheries Science Center, 8604 La Jolla Shores Drive, La Jolla, CA 92037.
- Santurtún, E., and F. Galindo. 2002. Coping behaviors of spotted dolphins during fishing sets. Administrative Report No. LJ-02-36C, NMFS, Southwest Fisheries Science Center, 8604 La Jolla Shores Drive, La Jolla, CA 92037.
- Southern, S., A. Allen, and N. Kellar. 2002. Molecular signature of physiological stress based on protein expression profiling of skin. Administrative Report No. LJ-02-27, NMFS, Southwest Fisheries Science Center, 8604 La Jolla Shores Drive, La Jolla, CA 92037.
- St Aubin, D. J. 2002. Further assessment of the potential for fishery-induced stress on dolphins in the eastern tropical Pacific. Administrative Report No. LJ-02-23C, NMFS, Southwest Fisheries Science Center, 8604 La Jolla Shores Drive, La Jolla, CA 92037.
- St. Aubin, D. J. 2002. Hematological and serum chemical constituents in pantropical spotted dolphins (*Stenella attenuata*) following chase and encirclement. Administrative Report No. LJ-02-37C, NMFS, Southwest Fisheries Science Center, 8604 La Jolla Shores Drive, La Jolla, CA 92037.

5. Quantitative stock assessment of depleted dolphins:

- Archer, F. and S. J. Chivers. 2002. Age structure of the Northeastern spotted dolphin incidental kill by year for 1971 to 1990 and 1996 to 2000. Administrative Report No. LJ-02-12, NMFS, Southwest Fisheries Science Center, 8604 La Jolla Shores Drive, La Jolla, CA 92037.**
- Archer, F., T. Gerrodette and A. Jackson. 2002. Preliminary estimates of the annual number of sets, number of dolphins chased, and number of dolphins captured by stock in the tuna purse-seine fishery in the eastern tropical Pacific, 1971-2000. Administrative Report No. LJ-02-10, NMFS, Southwest Fisheries Science Center, 8604 La Jolla Shores Drive, La Jolla, CA 92037.
- Chivers, S. J. 2002. Age structure of female eastern spinner dolphins (*Stenella longirostris orientalis*) incidentally killed in the eastern tropical Pacific tuna purse-seine fishery. Administrative Report No. LJ-02-11, NMFS, Southwest Fisheries Science Center, 8604 La Jolla Shores Drive, La Jolla, CA 92037.

Wade, P. R. 2002. Assessment of the population dynamics of the northeastern offshore spotted and the eastern spinner dolphin populations through 2002. Administrative Report No. LJ-02-13, NMFS, Southwest Fisheries Science Center, 8604 La Jolla Shores Drive, La Jolla, CA 92037.

6. Coastal spotted dolphins:

Escorza-Trevino, S., A. Lang and A. E. Dizon. 2002. Genetic differentiation and intraspecific structure of eastern tropical Pacific spotted dolphins, *Stenella attenuata*, revealed by mitochondrial and microsatellite DNA analyses. Administrative Report No. LJ-02-38, NMFS, Southwest Fisheries Science Center, 8604 La Jolla Shores Drive, La Jolla, CA 92037.